

# Automatic Road Recognition 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**—Road recognition is one of the most important steps in many computer vision applications, such as automatic driving, driver assistance systems, and traffic video surveillance. Various weather conditions and illumination effects cause the automatic road detection to be a challenging problem. In this paper, a new real-time road recognition method is proposed that is able to accurately extract the road region in traffic videos under adverse weather and illumination conditions. In particular, the novel global foreground modeling (GFM) method is applied to subtract the ever-changing background in the traffic video frames while robustly detecting the moving vehicles which are soundly assumed to pass along the road region. Initial road samples are obtained from the subtracted background model in the location of the moving vehicles. A combination of features from RGB, grayscale, and HSV color spaces are used to construct a probability map based on the standardized Euclidean distance between the feature vectors. The initially estimated road region is then integrated with the regions located by the flood-fill algorithm in order to obtain an accurate road mask. Experimental results using a dataset containing real traffic videos demonstrate the feasibility of the proposed method for automatic road recognition in real-time.

## I. INTRODUCTION

Road region extraction is a fundamental step in many modern computer vision applications, such as automatic driving, traffic warning, navigation, traffic surveillance, and driver assistance systems. Many studies have addressed the problem of vision-based road detection in recent years, in both applications of in-vehicle perception [5], [18], [21], [33], [46] and traffic video surveillance [11]–[13], [20], [29], [35]. The methods proposed in these studies are mostly applicable in both areas with some differences in the main motivations. Road detection helps with automatic driving, navigational warning and obstacle avoidance in the first group of applications and it is useful for region of interest (RoI) determination and traffic incidents detection in the second group.

The process of detecting the road in the case of applications like advanced driver assistance systems where the camera is usually mounted behind the car's windshield involves more illumination changes and variations in the road structure and therefore the performance of these methods tends to depend more on single-image road extraction. On the contrary, traffic surveillance videos captured by stationary cameras contain less rapid changes in the scene and therefore, temporal features are more reliable in the process of road detection. However, since the objects are closer to the camera in case of in-vehicle perception applications, color and texture features contain

more useful information compared to the surveillance videos where the camera sensor is usually placed in a further position overlooking the roadway.

Currently, in most traffic surveillance video applications the RoI is selected manually or estimated automatically by applying the accumulative foreground mask obtained during a number of video frames. However, these approaches have some limitations that result in a need for an automatic and adaptive road detection method.

- In manual selection of the RoI, the tiresome process is needed to be carried out separately for every different camera location.
- Utilizing the foreground mask for automatic determination of the RoI eliminates the need for manual labor; however, it depends entirely on the performance of the foreground segmentation method which in turn can suffer from challenging illumination conditions.
- The RoI determination method based on the accumulative foreground mask does not consider the features of each video frame and demands a number of vehicles to pass along different areas of the roadway in order to obtain a relatively good estimation of the RoI which can take an uncertain period of time.
- In Both strategies, not efficient information is provided regarding the road lanes and boundaries which can be helpful in many traffic analysis tasks, such as detecting vehicle accidents, stopped vehicles, and traffic congestion.
- Both approaches are required to be repeated from scratch every time the camera viewing angle or zoom settings changes.

This study is focused on automatic road region extraction in traffic videos that aids with RoI determination which in turn reduces the need for computational resources and can be useful in automated detection of traffic incidents and driving violations. The remainder of this paper is organized as follows. Section II outlines some of the recent studies that have approached the problem of road region extraction. Section III describes the proposed road recognition method in details. In section III-A a method for initial road recognition based on color differences is introduced. Section III-B contains details on refining the extracted road region by using temporal and color features. The performance of the proposed method is

evaluated in section IV on real traffic videos, and the paper is concluded in section V.

## II. RELATED WORK

The recent studies in road detection and RoI determination choose different strategies to segment the road region in the images. In some studies, the local features such as color [14], [24], [25], [34], [41], brightness [4], [40], [42], texture [19], [43], [45], or a combination of them [2], [15], [44] are extracted in order to classify the pixels into road and non-road classes. These group of methods are not sensitive to road shapes, but their performance can suffer from illumination effects. Some methods tend to rely on the road models in order to match them with low-level features and detect the road region [8], [16], [17], [26], [38]. These so called model-based methods are more robust in terms of dealing with different illumination conditions, but they are limited to a number of pre-defined road shape models and their performance can suffer in detecting unstructured roads. Several techniques suggest utilizing motion information and temporal features obtained from a sequence of video frames in order to extract and update the active traffic region which mostly represents the road area [20], [29], [39]. These group of methods are only applicable in cases where the input data is a video captured by a stationary camera and their performance relies on the background subtraction technique utilized to extract the location of moving objects.

Recently, convolutional deep neural networks have also been applied to segment the road region due to their ability in modeling non-linear variable relationships [1], [3], [7], [22], [23], [27], [30], [32]. The supervised techniques based on deep learning require large datasets which should be manually labeled which is a time-consuming and tedious task, especially in case of semantic segmentation where the annotators need to outline the objects carefully in each image. These group of methods also have high hardware requirements and due to generalization limitations they can only perform well if the unseen tested data is close to the data used to train the models. However, real-world applications require methods that are adaptable to various input data and can perform in real-time. In terms of road detection in traffic video analytic applications, the performance of supervised methods can suffer from a wide range of different illumination and weather conditions, image resolutions, camera's viewing angle, and distance from the road surface.

## III. A FEATURE-BASED STATISTICAL METHOD FOR ROAD REGION RECOGNITION

Determining the region of interest (RoI) is crucial pre-processing step in many video analytic applications. The focus of this study is on finding the RoI in traffic videos, which is associated with the roadway region, in real-time and with no manual input. In this section, the main steps of the proposed method are discussed in details, which has mainly two contributions: (i) The new sampling approach can approximate the roadway location can approximate the



Fig. 1: Sampling road regions from the background image based on the direction of moving vehicles. The red color indicated the sampled road pixels.

road location during the initial video frames solely based on color and temporal features without any assumption about the structure of the roadway. (ii) The novel road region extraction method can automatically determine the RoI to be used in traffic video applications.

### A. Initial road region approximation

In the tasks of traffic videos analysis, the RoI is associated with the roadway. In order to detect the location of the moving vehicles, a foreground segmentation method [36], [37] is applied which models the background locally while modeling the foreground globally. The road samples are taken from the subtracted background image in the corresponding location of the moving vehicles. An effective blob-tracking method [6] is utilized in order to estimate the moving direction of each vehicle. The foreground mask of each vehicle is then cropped based on its moving direction in order to filter out the potential non-road pixels. Figure 1 shows examples of the road sampling approach that is applied to construct a aggregated set of road samples  $\Omega_{sm}$ .

The sampled road pixels are used in order to generate a road probability map based on color differences. According to the histogram models of the road samples, an example of which can be seen in Figure 2, the blue, green, and grayscale values of each pixel are proven to be discriminative features for classifying the road and non-road regions. A set of four-dimensional feature vectors is denoted by:

$$\mathcal{F}^t = \{f_1^t, f_2^t, \dots, f_N^t\} = \{f_i^t\}_{i=1}^N \quad (1)$$

where  $f_i^t$  is a  $D$  dimensional feature vector of pixel  $i$  that contains the blue, green, grayscale, and hue values of that pixel at frame  $t$ . The standardized Euclidean distance between each feature vector and the mean value of the pixels in the road sample are calculated as follows:

$$\begin{aligned} \bar{f}_j^t &= \left( \sum_{i \in \Omega_{sm}} f_{ij}^t \right) / |\Omega_{sm}^t| , j = 1 \dots D \\ (\sigma_j^2)^t &= \left( \sum_{i \in \Omega_{sm}} (f_{ij}^t - \bar{f}_j^t)^2 \right) / |\Omega_{sm}^t| , j = 1 \dots D \\ d_i^t &= \sqrt{\sum_{j=1}^D \frac{1}{(\sigma_j^2)^t} (f_{ij}^t - \bar{f}_j^t)^2} \end{aligned} \quad (2)$$

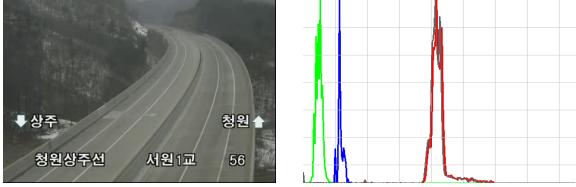


Fig. 2: The histogram plot representing the RGB and gray values of the background image from the road samples.

where  $i = 1 \dots N$  is the pixel index,  $D = 4$  is the number of features,  $\bar{f}_j^t$  is the mean value of the  $j$ -th feature in the set of road samples  $\Omega_{sm}$ ,  $f_{ij}^t$  is the  $j$ -th feature of pixel  $i$ ,  $d_i^t$  is the standardized Euclidean distance,  $(\sigma_j^t)^2$  is the variance of the  $j$ -th feature at frame  $t$ . The road probability map  $\mathcal{P}_{\mathcal{R}}$  is denoted by:

$$\mathcal{P}_{\mathcal{R}}^t = \{p_1^t, p_2^t, \dots, p_N^t\} = \{p_i^t\}_{i=1}^N \quad (3)$$

where  $N$  is the total number of pixels, and  $p_i \in [0, 1]$  is the road probability value of pixel  $i$  at frame  $t$  which is calculated as follows:

$$\begin{aligned} \sigma_j^t &= \sqrt{\left( \sum_{i \in \Omega_{sm}} (f_{ij}^t - \bar{f}_j^t)^2 \right) / |\Omega_{sm}^t|}, j = 1 \dots D \\ \bar{\sigma}^t &= \left( \sum_{j=1}^D \sigma_j^t \right) / D \\ \lambda_i^t &= \max \left( 0, \text{sgn} \left( d_i^t - \bar{\sigma}^t \right) \right) \\ p_i^t &= 1 - \lambda_i^t \left( \frac{d_i^t}{\sigma^t} + \frac{1}{k^2} \right), k - 1 \leq \frac{d_i^t}{\sigma^t} < k \end{aligned} \quad (4)$$

where  $i = 1 \dots N$  is the pixel index,  $\bar{\sigma}^t$  is the mean standard deviation of the features among the  $D$  dimensions in  $\Omega_{sm}$ ,  $k$  is a natural number in  $\{k \in \mathbb{N} | 1 < k \leq \max(d_i^t - \bar{\sigma}^t)\}$ , and  $p_i$  is the resulting probability value of pixel  $i$  at frame  $t$ . Figure 3 represents an example of the extracted probability map from the difference values.

The extracted probability map  $\mathcal{P}_{\mathcal{R}}$  is updated throughout the video frames by applying the temporal fusing algorithm as follows:

$$\begin{aligned} \hat{p}_i^t &= \frac{\sum_{f=1}^t w^f p_i^f}{1 + \sum_{f=1}^t w^f} \\ w^f &= |\Omega_{sm}^f| \end{aligned} \quad (5)$$

where  $i = 1 \dots N$  is the pixel index,  $w^f$  is the weight of frame  $f$  which is associated with the number of pixels in the aggregated sample mask,  $p_i^f$  is the probability value of pixel  $i$  at frame  $f$ ,  $\Omega_{sm}^f$  is the accumulative road sample mask at frame  $f$ ,  $N$  is the total number of pixels in each frame, and  $\hat{p}_i^t$  is the updated probability value of pixel  $i$ . The Otsu's threshold [10] is applied on the resulting map in order to obtain a binary mask  $\mathcal{P}_{\mathcal{R}}^*$ .

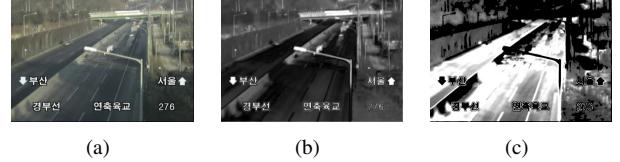


Fig. 3: Extracting the auxiliary road region probability map using feature difference values. (a) The subtracted background image  $\mathcal{B}$ . (b) The difference image  $\mathcal{B}_d$  obtained by  $d_i$ . (c) The extracted probability map  $\mathcal{P}_{\mathcal{R}}$ .

#### B. Integration of temporal features with the estimated road region

Since road is a unified object, a combination of color and temporal features represent a more reliable estimation when enough time has passed. Here, the flood-fill algorithm is applied in order to unify the connected components of the road pixels and extract the road region. In order to define the limiting boundaries of the flood-fill method, the Canny edge detection method is applied on the difference image  $\mathcal{B}_d$  with  $\tau_l = 0.66 \times M$  and  $\tau_h = 1.33 \times M$  as the lower and upper thresholds, respectively, where  $M$  is the median luminance of  $\mathcal{B}_d$ . After applying the edge detection method, leak segmentation error can still occur due to lack of enough gradient information at the dominant road boundaries, which can be corrected by using the accumulative foreground mask  $\mathcal{F}_{acc}$ . Algorithm 1 shows the steps of accumulating the foreground masks obtained by the GFM [36], [37] method with false positives and slow-moving object filtered out by applying two thresholds  $\tau_d$  and  $\tau_s$  at steps 3–8. The threshold  $\tau_i$  is used to define how long a track has to be inactive before being removed. The accumulative foreground mask  $\mathcal{F}_{acc}$  is added by  $d$  in the location of the track only after track  $t$  has been removed from the set  $T$  (step 11 of Algorithm 1). This way the tracks with larger movements contribute more to the estimated road region. At the end,  $\mathcal{F}_{acc}$  is normalized as it is divided by the maximum value.

Similar to the approach used in [9], [28], the contours of  $\mathcal{F}_{acc}$  are smoothed using a Gaussian kernel. The Gaussian coefficients are calculated as follows:

$$\begin{aligned} \sigma' &= \frac{1}{2} (c\sigma + 1) \\ \mathcal{M} &= 2 (\text{sgn}(\sigma') \lfloor |\sigma'| + 0.5 \rfloor) - 1 \\ g_i &= \alpha \exp \left( - \frac{(i - \frac{\mathcal{M}-1}{2})^2}{2\sigma^2} \right), \sum_{i=0}^{\mathcal{M}-1} g_i = 1 \end{aligned} \quad (6)$$

where  $c$  is an integer constant,  $\mathcal{M} \in \{2n + 1 : n \in \mathbb{Z}\}$  is the Gaussian aperture size,  $\sigma$  is the standard deviation,  $\alpha$  is the scale factor chosen so that  $\sum_{i=0}^{\mathcal{M}-1} g_i = 1$ , and  $g_i$  is the  $i$ -th Gaussian filter coefficient. The contours are smoothed

---

**Algorithm 1:** Acquiring the accumulative foreground mask

---

**Input:**

The size of each video frame  
 The set  $T$  of vehicle tracks in the current frame  
 The set of blobs for each track  $B_t = \{b_1, \dots, b_n\}$   
 A set of predefined thresholds  $\mathcal{T} = \{\tau_d, \tau_i, \tau_s\}$

**Output:**

The accumulative foreground mask  $\mathcal{F}_{acc}$  of the same size as the video frame

```

1 initialize  $\mathcal{F}_{acc}$  with 0;
2 foreach  $t \in T$  do
3   if  $size(t) < \tau_s$  then
4     | continue;
5   end
6    $d = \|t_{cn} - t_{c1}\|$ ;
7   if  $d < \tau_d$  then
8     | continue;
9   end
10  add track's current blob  $b_n$  to track's accumulative
      mask  $\mathcal{F}_t$ ;
11  if  $t_i > \tau_i$  then
12    |  $\mathcal{F}_{acc}[\mathcal{F}_t] = \mathcal{F}_{acc}[\mathcal{F}_t] + d$ ;
13  end
14 end
15  $\mathcal{F}_{acc} = \frac{\mathcal{F}_{acc}}{\max(\mathcal{F}_{acc})}$ ;

```

---

separately over each  $X$  and  $Y$  axis:

$$C_j^k(n) = \begin{cases} C(|C| + n - k) & , \text{if } n < k \\ C(n - k - |C|) & , \text{if } n > (k + |C| - 1) \\ C(n - k) & , \text{otherwise} \end{cases} \quad (7)$$

$$C_j^*(n) = \sum_{i=0}^{M-1} C_j^k(n) g_i \quad , \quad k = -\mathcal{L} \dots \mathcal{L}$$

where  $n = 0 \dots (|C| - 1)$  is the index of each point on the curve,  $C$  is the surrounding contour of the accumulative foreground mask,  $j \in \{x, y\}$  represents the  $x$  or  $y$  axis,  $\mathcal{L} = \frac{1}{2}(\mathcal{M} - 1)$ , and  $C_j^*(n)$  is the position of the  $n$ -th point in the smoothed contour.

The sides of the smoothed contours which correspond to the boundaries of the road are partitioned into a set of  $K$  separate clusters  $\mathcal{C} = \{c_k\}_{k=1}^K$  based on their connectivity which is in turn measured by Euclidean distance. The points of each cluster  $c_k$  are resampled by traversing in a pace equal to resample size  $m_k = s_k/d$  where  $s_k$  is the arc-length of  $c_k$  and  $d$  is a pre-defined constant.

Then a similar approach to [31] is used to estimate the boundaries of the road by fitting a second-degree polynomial curve on each cluster. The principal component analysis (PCA) method is applied on each set of re-sampled points in order to calculate the direction of the maximum variation in the set. First a matrix  $\mathbf{P}_k \in \mathbb{N}^{m_k \times 2}$  is formed with each row containing the  $x, y$  coordinate values of each resampled point

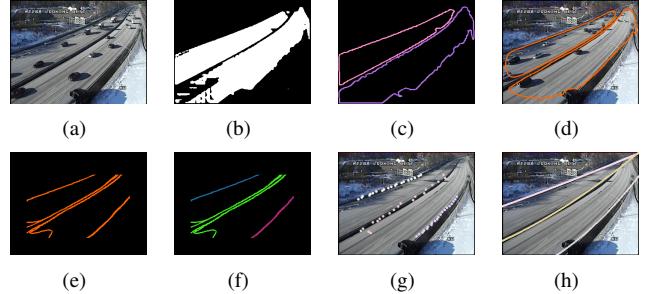


Fig. 4: Extracting the road region using the cumulative foreground masks of the moving vehicles. (a) The original traffic video frame. (b) Accumulative foreground mask after one minute. (c) Contours of the accumulative foreground masks. (d) Smoothed contours. (e) Cropped contours. (f) Clustering. (g) Resampled points. (h) Estimated road boundaries.

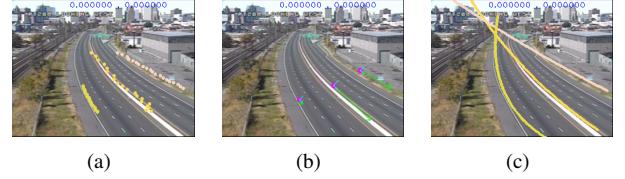


Fig. 5: Extracting the dominant road boundaries using the PCA method. (a) Re-sampled points used for curve fitting. (b) The direction of the maximum variation recognized by PCA. (c) The limiting boundaries estimated by curve fitting.

from  $c_k$ . Then the covariance matrix  $\mathbf{S}_k$  is computed as follows:

$$\mathbf{u}_k = \frac{1}{m_k} \sum_{i=1}^{m_k} \mathbf{P}_k \quad (8)$$

$$\mathbf{S}_k = \frac{1}{m_k - 1} \sum_{i=1}^{m_k} (\mathbf{P}_k - \mathbf{u}_k)(\mathbf{P}_k - \mathbf{u}_k)^T$$

where  $\mathbf{u}_k$  is a row vector that contains the mean  $\bar{x}$  and  $\bar{y}$  values of each column in  $\mathbf{P}_k$ . The eigenvalues and eigenvectors of the covariance matrix are calculated as follows:

$$\lambda_1^k, \lambda_2^k = \frac{1}{2} \left( \sigma_{x_k}^2 + \sigma_{y_k}^2 \pm \sqrt{(\sigma_{x_k}^2 - \sigma_{y_k}^2)^2 + 4\sigma_{x_k y_k}^2} \right) \quad (9)$$

$$\mathbf{e}_j^k = \frac{1}{\sqrt{\sigma_{x_k y_k}^2 + (\lambda_j^k - \sigma_{x_k}^2)^2}} \begin{bmatrix} \sigma_{x_k y_k}^2 \\ \lambda_j^k - \sigma_{x_k}^2 \end{bmatrix}$$

where  $j \in \{1, 2\}$ ,  $\sigma_{x_k}^2$ ,  $\sigma_{y_k}^2$ , and  $\sigma_{x_k y_k}^2$  are the variance of  $x$ , variance of  $y$ , and covariance of  $xy$  values in  $\mathbf{P}_k$ , respectively.  $\lambda_j^k$  and  $\mathbf{e}_j^k$  are the eigenvalues and their corresponding eigenvectors of  $\mathbf{S}_k$ . A matrix  $\mathbf{E}_k$  is defined as follows:

$$\mathbf{E}_k = \begin{bmatrix} a_{11}^k & a_{12}^k \\ a_{21}^k & a_{22}^k \end{bmatrix} \quad (10)$$

where  $\mathbf{e}_1^k = [a_{11}^k, a_{21}^k]^T$  and  $\mathbf{e}_2^k = [a_{12}^k, a_{22}^k]^T$  are the first and second eigenvectors of  $\mathbf{P}_k$ , respectively. A new axis is

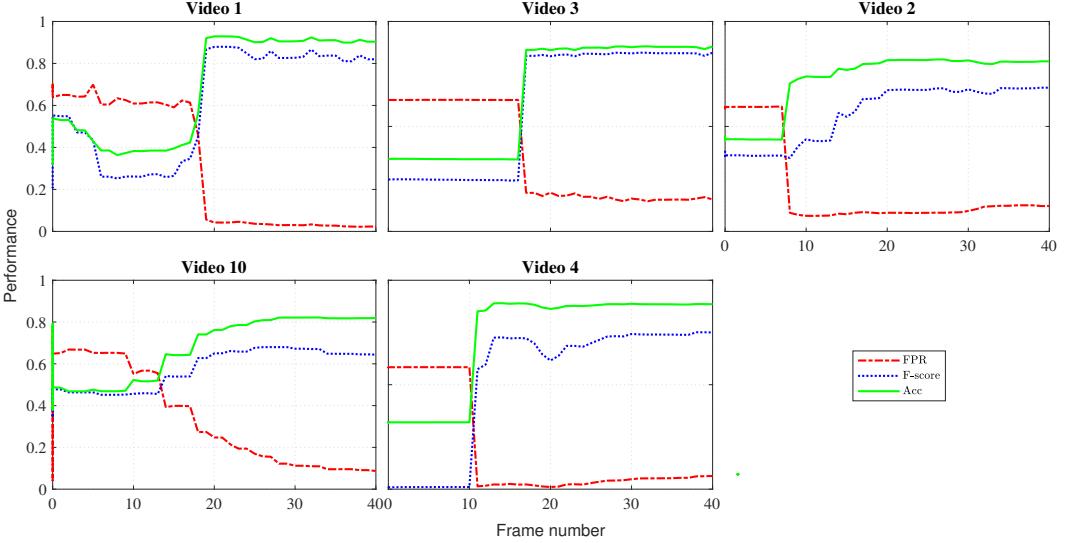


Fig. 6: The F-measure score, accuracy, and false-positive rate of the proposed road extraction method at different frames, tested on some of the sample traffic videos. The sudden improvement in the performance measures happens when the first vehicle is observed in the video sequence and the initial road samples are obtained based on its location. The accuracy and F1-score keep improving to more than 90% in most videos.

generated and the data points from  $\mathbf{P}_k$  are rotated as follows:

$$\begin{aligned}\theta_k &= \cos^{-1}(tr(\mathbf{E}_k)/2) \\ \mathbf{R}_k &= \begin{bmatrix} \cos \theta_k & -\sin \theta_k \\ \sin \theta_k & \cos \theta_k \end{bmatrix} \\ \mathbf{P}'_k^T &= \mathbf{R}_k \mathbf{P}_k^T\end{aligned}\quad (11)$$

where  $\theta_k$  is the direction of maximum dispersion in  $\mathbf{P}_k$ ,  $tr(\mathbf{E}_k) = a_{11}^k + a_{22}^k$ ,  $\mathbf{R}_k$  is the rotation matrix, and  $\mathbf{P}'_k$  is the matrix containing the rotated points. After second-degree polynomial curve-fitting on each  $\mathbf{P}'_k$ , the resulting curves are rotated back to the original  $x$  and  $y$  axis to represent an estimation of the dominant road boundaries. Figures 4 and 5 represent an example of road boundary estimation.

The flood-fill algorithm with a connectivity value of 4 and seed points are taken from the set  $\Omega_{sm}$  is applied in order to aggregate the connected components in a mask image  $\mathcal{M}_{\mathcal{F}}$ . During the component connection process, the maximal lower and upper thresholds of intensity difference between the current pixel and each of its nearest neighbors of the same component, or a new seed pixel being added to the component are calculated based on the standard deviation of the background image  $\mathcal{B}$  as follows:

$$\begin{aligned}\bar{\mathcal{B}} &= \frac{1}{N} \sum_{i=1}^N \mathcal{B}_i \\ \sigma &= \sqrt{\frac{\sum_{i=1}^N (\mathcal{B}_i - \bar{\mathcal{B}})^2}{N}} \\ \tau &= \max(1, \frac{\sigma}{k})\end{aligned}\quad (12)$$

where  $\bar{\mathcal{B}}$  is the mean value of the background image  $\mathcal{B}$ ,  $N$  is the total number of pixels in the background image,  $\mathcal{B}_i$  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.

When the intersection between the binary probability mask  $\mathcal{P}_{\mathcal{R}}^*$  and the aggregated flood-fill mask  $\mathcal{M}_{\mathcal{F}}$  surpasses a threshold, the flood-fill algorithm has connected most of the road components. Morphological procedure is performed on  $\mathcal{M}_{\mathcal{F}}$  to bridge the gaps and the intersection between its result and  $\mathcal{P}_{\mathcal{R}}^*$  is utilized as the final estimated road region as follows:

$$\begin{aligned}\mathcal{M}'_{\mathcal{F}} &= \mathcal{M}_{\mathcal{F}} \oplus B \\ \mathcal{T} &= \frac{|\mathcal{M}'_{\mathcal{F}} \cap \mathcal{P}_{\mathcal{R}}^*|}{|\mathcal{P}_{\mathcal{R}}^*|} \\ \mathcal{M}_{\mathcal{R}} &= \begin{cases} \mathcal{P}_{\mathcal{R}}^* & , \text{if } \mathcal{T} < \theta \\ \mathcal{M}'_{\mathcal{F}} \cap \mathcal{P}_{\mathcal{R}}^* & , \text{otherwise} \end{cases}\end{aligned}\quad (13)$$

where  $\mathcal{M}'_{\mathcal{F}} = \{x | [(\hat{\mathcal{B}})_x \cap \mathcal{M}_{\mathcal{F}}] \neq \emptyset\}$  is the result of a dilation operation with  $B$  as a structuring element,  $\mathcal{T}$  is the number of common pixels between the probability mask and the accumulative flood-fill mask,  $\theta \in [0, 1]$  is a predefined threshold, and  $M_R$  is the final mask representing road pixels.

#### IV. EXPERIMENTS

The performance of the proposed method is evaluated using several videos. The dataset is provided by the New Jersey Department of Transportation (NJDOT) which contains real traffic videos captured from highways with different road structures, various illumination conditions, resolutions and frame-rates. A single frame from each tested video is shown at the first rows in Figures 7 and 8. The second rows represent the ground-truth mask corresponding to the road region in each video. The last rows illustrate the extracted road by a red-colored mask over the subtracted background of each video

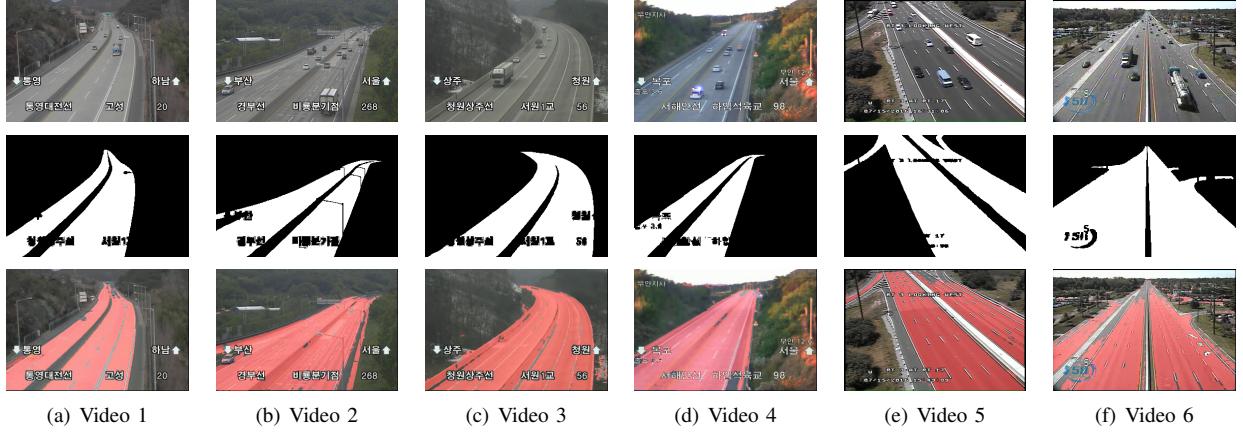


Fig. 7: Road extraction results in regular traffic videos. The first row displays a sample frame of each video. The second row represents the ground-truth road region masks. The third row illustrates the extracted road region by the proposed method before applying the accumulative foreground mask.

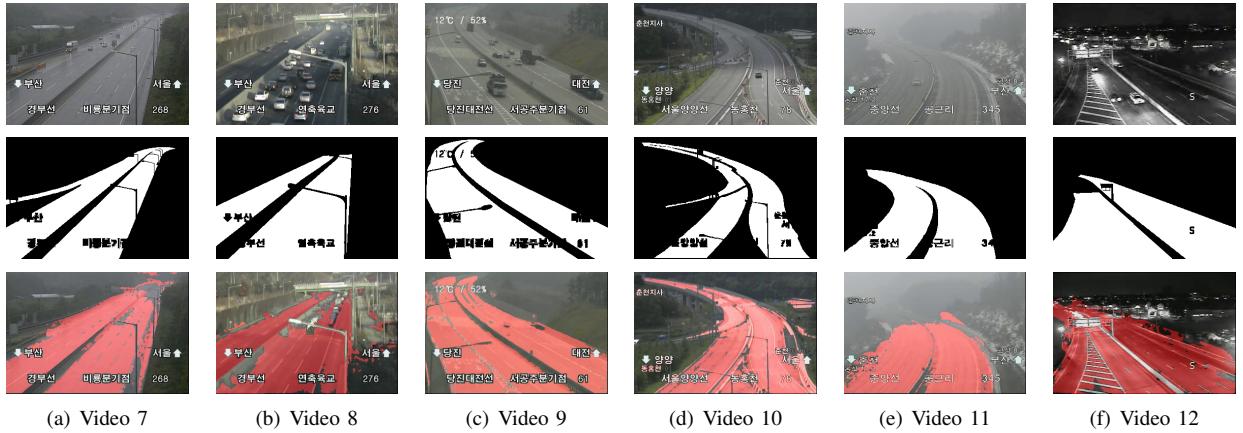


Fig. 8: Road extraction results in traffic videos with challenging illumination conditions. The first row displays a sample frame of each video. The second row represents the ground-truth road region masks. The third row illustrates the extracted road region by the proposed method before applying the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

frame. All the experiments were conducted using a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. The average processing speed for video frames of size  $720 \times 480$  pixels was around  $\sim 42.22$  frames per second, which is inline with real-time requirements of video analysis applications.

The following metrics are used in order to evaluate the quantitative results:

$$\begin{cases} FPR = F_P / (F_P + T_N) \\ PRE = T_P / (T_P + F_P) \\ REC = T_P / (T_P + F_N) \\ F_1 = 2 \times (PRE \times REC) / (PRE + REC) \\ ACC = (T_P + T_N) / (T_P + F_P + T_N + F_N) \end{cases} \quad (14)$$

where  $T_P$ ,  $F_P$  are the number of pixels correctly and incorrectly reported as road regions, and  $T_N$  and  $F_N$  are the number of pixels that are correctly and incorrectly reported as non-

road regions, respectively.  $FPR$ ,  $PRE$ ,  $REC$ ,  $ACC$ , and  $F_1$  refer to false positive rate, precision, recall, and F1-score, and accuracy, respectively. Figure 6 shows the changes in F1-score, accuracy, and the false-positive rate based on the frame number for some traffic videos.

When the first vehicle is observed in each video, the initial road samples are obtained from its location and detection results improve when more vehicles pass along the road. In Table I the quantitative performance of the road region extraction method is reported for the 12 sample traffic videos. Due to under-segmentation, the entire roadway region is not always extracted which results in the precision values being slightly higher than the recall values in most cases. Examples of under-segmentation can be seen in Figures 7(e), 7(f), 8(a) and 8(e) which is mostly caused by the loss of tracking information at the far side of the road. Strong cast shadows and traffic congestion can also result in under-segmentation, specially at

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

the initial frames (e.g., Figure 8(b)). Over-segmentation or leak segmentation can also occur which is usually due to the lack of sufficient gradient information at the road boundaries or the illumination effects which causes similarity between the non-road and road regions. Examples of over-segmentation are exemplified in Figures 7(b), 7(d), 8(e) and 8(f).

## V. CONCLUSION

Determining the region of interest (RoI) in video analysis applications is an important pre-processing step. An accurate RoI can keep all the essential information for the video analysis tasks, while reducing the need for computational resources and eliminating a portion of potential false outputs. In applications of traffic video analysis the RoI usually refers to the road region. In this paper, an adaptive and fully automatic method is proposed for road recognition in real-time. Temporal features are utilized in a robust statistical foreground segmentation method in order to obtain sample data based on the location of the moving vehicles. The road region is then extracted in the initial frames of the video by using pixel values in different color-spaces. The initially extracted road is updated and refined with more sample data generated through the subsequent video frames.

The proposed method shows good performance in videos of different sizes and frame-rates with various illumination and weather conditions. This method is not limited to a number of geometrical models since no assumptions have been made about the structure or shape of the road. The extracted road location can further be utilized as the RoI in different traffic video analysis tasks. The experimental results from evaluating the method using real traffic videos provided by NJDOT indicate the feasibility of the proposed method for real-world applications.

## ACKNOWLEDGMENT

This paper is partially supported by the NSF grant 1647170.

## REFERENCES

- [1] 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.
- [2] B. Benligiray, C. Topal, and C. Akinlar, “Video-based lane detection using a fast vanishing point estimation method,” in *2012 IEEE International Symposium on Multimedia*. IEEE, 2012, pp. 348–351.
- [3] L. Caltagirone, L. Svensson, M. Wahde, and M. Sanfridson, “Lidar-camera co-training for semi-supervised road detection,” *arXiv preprint arXiv:1911.12597*, 2019.
- [4] A. F. Cela, L. M. Bergasa, F. L. Sanchez, and M. A. Herrera, “Lanes detection based on unsupervised and adaptive classifier,” in *2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks*. IEEE, 2013, pp. 228–233.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] G. Cheng, Y. Wang, Y. Qian, and J. H. Elder, “Geometry-guided adaptation for road segmentation,” in *2020 17th Conference on Computer and Robot Vision (CRV)*. IEEE, 2020, pp. 46–53.
- [9] S. Elghoul and F. Ghorbel, “An efficient 2d curve matching algorithm under affine transformations,” in *VISIGRAPP (4: VISAPP)*, 2018, pp. 474–480.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, “Automatic traffic surveillance system for vehicle tracking and classification,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 175–187, 2006.
- [14] J. Huang, B. Kong, B. Li, and F. Zheng, “A new method of unstructured road detection based on hsv color space and road features,” in *2007 International Conference on Information Acquisition*. IEEE, 2007, pp. 596–601.
- [15] T. Huang, Z. Wang, X. Dai, D. Huang, and H. Su, “Unstructured lane identification based on hough transform and improved region growing,” in *2019 Chinese Control Conference (CCC)*. IEEE, 2019, pp. 7612–7617.
- [16] S.-N. Kang, S. Lee, J. Hur, and S.-W. Seo, “Multi-lane detection based on accurate geometric lane estimation in highway scenarios,” in *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, 2014, pp. 221–226.
- [17] H. U. Khan, A. R. Ali, A. Hassan, A. Ali, W. Kazmi, and A. Zaheer, “Lane detection using lane boundary marker network with road geometry constraints,” in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 1834–1843.
- [18] 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.
- [19] H. Kong, J.-Y. Audibert, and J. Ponce, “General road detection from a single image,” *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2211–2220, 2010.
- [20] 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.
- [21] T. Kühnl, F. Kummert, and J. Fritsch, “Monocular road segmentation using slow feature analysis,” in *2011 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2011, pp. 800–806.
- [22] A. Laddha, M. K. Kocamaz, L. E. Navarro-Serment, and M. Hebert, “Map-supervised road detection,” in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2016, pp. 118–123.
- [23] 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.
- [24] D. Liu, Y. Wang, and T. Chen, “Application of color filter adjustment and k-means clustering method in lane detection for self-driving cars,” in *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2019, pp. 153–158.
- [25] D. Liu, Y. Wang, T. Chen, and E. T. Matson, “Accurate lane detection for self-driving cars: An approach based on color filter adjustment and k-means clustering filter,” *International Journal of Semantic Computing*, vol. 14, no. 01, pp. 153–168, 2020.
- [26] S. Luo, X. Zhang, J. Hu, and J. Xu, “Multiple lane detection via combining complementary structural constraints,” *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [27] 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.
- [28] F. Mai, C. Chang, and Y. Hung, “Affine-invariant shape matching and recognition under partial occlusion,” in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 4605–4608.
- [29] J. Melo, A. Naffet, 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.
- [30] G. Oliveira, W. Burgard, and T. Brox, “Efficient deep methods for monocular road segmentation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016)*, 2016.
- [31] N. Parvin, “Robust curved road boundary identification using hierarchical clustering.” Ph.D. dissertation, 2013.
- [32] 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.
- [33] T. Rateke and A. von Wangenheim, “Road surface detection and differentiation considering surface damages,” *arXiv preprint arXiv:2006.13377*, 2020.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] ——, “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.
- [38] S. Srivastava and R. Maiti, “Multi-lane detection robust to complex illumination variations and noise sources,” in *2019 1st International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*. IEEE, 2019, pp. 1–8.
- [39] L.-W. Tsai, Y.-C. Chean, C.-P. Ho, H.-Z. Gu, and S.-Y. Lee, “Multi-lane detection and road traffic congestion classification for intelligent transportation system,” *Energy Procedia*, vol. 13, pp. 3174–3182, 2011.
- [40] T. Veit, J.-P. Tarel, P. Nicolle, and P. Charbonnier, “Evaluation of road marking feature extraction,” in *2008 11th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2008, pp. 174–181.
- [41] B. Wang, V. Frémont, and S. A. Rodríguez, “Color-based road detection and its evaluation on the kitti road benchmark,” in *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, 2014, pp. 31–36.
- [42] J. Wang, T. Mei, B. Kong, and H. Wei, “An approach of lane detection based on inverse perspective mapping,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 35–38.
- [43] X. Wang, Y. Guo, J. Wang, and M. Song, “Self-made texture and clustering based road recognition for ugv,” in *2017 36th Chinese Control Conference (CCC)*. IEEE, 2017, pp. 10999–11004.
- [44] L. Xiao, R. Wang, B. Dai, Y. Fang, D. Liu, and T. Wu, “Hybrid conditional random field based camera-lidar fusion for road detection,” *Information Sciences*, vol. 432, pp. 543–558, 2018.
- [45] G. Yang, Y. Wang, J. Yang, and Z. Lu, “Fast and robust vanishing point detection using contourlet texture detector for unstructured road,” *IEEE Access*, vol. 7, pp. 139 358–139 367, 2019.
- [46] L. Zhang and E.-y. Wu, “A road segmentation and road type identification approach based on new-type histogram calculation,” in *2009 2nd International Congress on Image and Signal Processing*. IEEE, 2009, pp. 1–5.