A New Adaptive Bidirectional Region-of-Interest Detection Method for Intelligent Traffic Video Analysis

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—**Real-time intelligent video-based traffic surveillance applications play an important role in intelligent transportation systems. To reduce false alarms as well as to increase computational efficiency, robust road segmentation for automated Region of Interest (RoI) detection becomes a popular focus in the research community. A novel Adaptive Bidirectional Detection (ABD) of region-of-interest method is presented in this paper to automatically segment the roads with bidirectional traffic flows into two regions of interest. Specifically, a foreground segmentation method is first applied along with the flood-fill algorithm to estimate the road regions. Then the Lucas-Kanade's optical flow algorithm is utilized to track and divide the estimated road into regions of interest in real-time. Experimental results using a dataset of real traffic videos illustrate the feasibility of the proposed method for automatically determining the RoIs in real-time.**

I. Introduction

A region of interest (RoI), is a sample within a dataset identified for a particular purpose [3]. In case of video analysis, a region of interest refers to a subspace of the video frame that is identified as the region of main focus. Selecting one or multiple regions of the video frame to perform video analytic tasks not only reduces the unnecessary and false results, but decreases the computational complexity due to a lower volume of input data which means a great deal to real-time applications. One of the main applications of video analysis is in traffic surveillance videos where the region of interest usually refers to the road area and its proximity. The area of focus in traffic video analysis tasks such as vehicle counting, speed estimation, and detecting traffic incidents such as wrong-way vehicles and vehicle accidents is the road lanes and shoulders. Currently, in most applications the region of interest is selected manually which has to be performed for every video and repeated in case of changes in the angle or distance of camera view.

Automatic road recognition has been a popular research topic in applications regarding traffic surveillance videos [10], [13], [18], [19] and in-vehicle perception [5], [12], [14], [23], [29]. Most of the techniques used in these studies are applicable in both areas with the main motivation of the former being RoI determination and the later providing useful information for advanced driving assistance systems. The focus of this study is road recognition and RoI determination in traffic surveillance videos to aid with detection of driving violations, traffic incidents recognition, and reducing the computational complexity of urban and highway traffic video analysis tasks. While some methods attempt to deal with road detection by extracting low-level and high-level features from single images and classify the pixels into road and non-road sets [10], [19], [20], [24], [25], [27], [28], other methods tend to utilize the motion information gained from a sequence of video frames in order to segment each frame into active and inactive traffic regions [4], [13], [15], [18].

In this paper we propose a motion-based statistical method to extract the road region and separate the road map into left and right sides based on the two major moving directions of vehicles in traffic videos. First, we apply a foreground detection technique based on Gaussian mixture models to detect the moving vehicles in a stable background. The color values of the background pixels at the corresponding vehicle locations are utilized as seed points by the flood-fill method in an accumulative manner to obtain an approximate region representing the road. The straight and curved road boundaries are estimated by second-degree polynomial curve-fitting to improve the obtained road map from the previous step through removing possible extra pixels that are incorrectly categorized as road pixels by the flood-fill method. At the same time, a statistical approach is applied with Lucas-Kanade optical flow and is further refined by a blob-tracking method to separate the two major directions in roads with bidirectional traffic. The detected road regions can further be updated and used as RoI in traffic video surveillance applications.

This paper is organized as follows. In section II, we will outline the previous related work that have approached the problem from various angles of view. In section III the proposed motion-based RoI detection method is described in detail. Section III-A describes how the entire road region is extracted using a number of consecutive video frames. Section III-B contains details on the proposed method for separating the segmented road region into two RoIs corresponding to the major directions in bidirectional traffic. The performance of the proposed method is evaluated on different traffic videos in section IV, and we conclude the paper in section V.

II. Related work

Automatic Region of Interest (RoI) detection is an important task in many traffic video analysis applications and can be used in road management, driver assistance systems, automatic driving, intelligent traffic surveillance, robot and car navigation systems, etc. In recent years, many automatic RoI detection methods have been proposed in order to reduce the manual work in urban and highway traffic monitoring applications. Some methods have tried to utilize various features in order to segmenting the road region from the remaining of an image. In [19], a feature vector of gray-amount, texture homogeneity, traffic motion and horizontal line is fed to support vector machine to classify each superpixel into road or non-road. Helala et al. [10] use the contours of superpixel blocks to generate a large number of edges which are organized into clusters of co-linearly similar sets and the best clusters are chosen according to a confidence level assigned to each cluster. At the end, the top-ranked pair of clusters are selected as road boundaries. Almazan et al. [1] combine a spatial prior with the vanishing point and horizontal line estimators in order to adapt to new weather conditions. In [7], a road segmentation method is proposed by applying Gaussian mixture model on color features and fusing it geometric cues within a Bayesian framework.

Some studies approach the task of roadway detection by using temporal features and extracting the active traffic regions. In [15], moving parts of the scene in videos of bidirectional traffic are extracted as difference images between two consecutive frames and accumulated to form a road map. Then a center line is used to divide the roadway into two regions each of which corresponding to one of the two major traffic directions. Similarly, Tsai et al. [26] accumulate the difference between two consecutive frames to obtain a map of the road where the motion vectors are used to separate the roadway into two regions in order to represent two major traffic directions. The performance of background subtraction and tracking methods utilized in this type of methods has a large influence on the results of the road segmentation process.

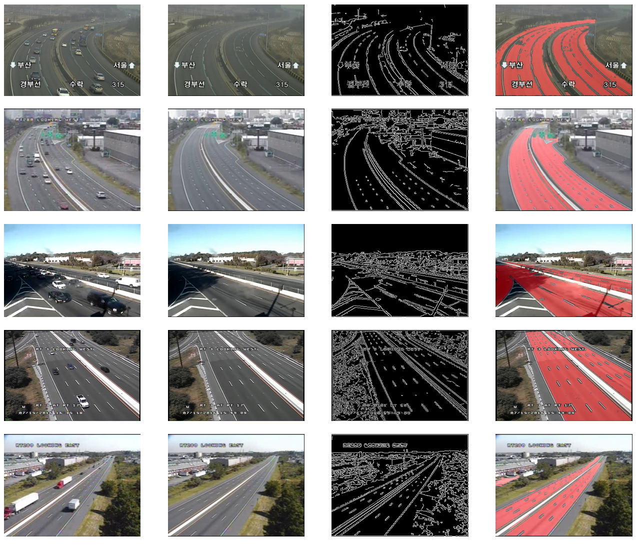
Most recent studies tend to propose illumination-invariant methods to deal with strong shadows and benefiting from the recent advances in deep learning models to segment the road in a supervised manner. Li et al. [16] propose a bidirectional fusion network (BiFNet) consisting of a dense space transformation module and context-base feature fusion module in order to fuse the image and the bird's eye view of the point cloud. In [25], an effective projection angle is calculated in logarithmic domain to extract the intrinsic images with weakened shadow effect and adopt to different directions of the camera view. Li et al. [17] propose a road segmentation by estimating the spatial structure of the road and using the color and edge features of the intrinsic image which is extracted based on regression analysis. Cheng et al. [8] propose a novel adaptation method to generalize road segmentation to new illumination situations and viewing geometries by training a fully-convolutional network for road segmentation. The learned geometric prior is anchored by estimating the vanishing point of the road and is used to extract road regions which are utilized as ground-truth data to adapt the network to the target domain. Wang et al. [27] generate an illumination invariant image and a manual triangular area is used as the color sample to obtain a number of probability maps which are used to segment the road which is further refined by the taking the extracted road boundaries into consideration. In [11], multiple abstract features from the explicitly derived representations of the video frames are extracted and fed to a shallow convolutional neural network. Most of the new studies benefit from supervised learning methods which limits their ability to adapt to new videos. Here, we proposed an unsupervised statistical method which can be applied in real-time applications.

III. An automatic motion-based method for extracting the RoI

Extracting the region of interest is an important pre-processing step in many image and video analytic applications. Currently, the selection of RoI is mostly performed manually by a human agent at initial stages of pre-processing. Retrieving the RoI automatically can reduce the need for manual work and constant updates in the extracted RoI helps with adaptation to new scenes when the camera's view changes. We propose a fully automatic method for road recognition which updates the RoI at each frame of the video and therefore can quickly adjust to changes in camera's view. Our proposed method has three major contributions: (i) The new motion-based statistical method can automatically extract the road region and reduce a great deal of manual work. (ii) The novel RoI determination approach can extract a separate RoI for each side of roads with bidirectional traffic. (iii) The RoI determination is fast and robust for real world application use.

A. A new method for road recognition based on Flood-Fill algorithm

To obtain an estimate of the road region in traffic videos captured by a stationary camera we can use the location of moving vehicles due to the fact that most vehicles pass along the road. In order to detect the moving vehicles we choose to use a statistical global foreground modeling (GFM) method based on mixture of Gaussians followed by Bayes classifier to segment the moving vehicles from static background [21], [22]. In the GFM method all the foreground pixels are modeled globally and the parameters are updated throughout the video which makes the approach adapt to different foreground objects. This method is also capable of continuously detect the vehicles even after they have stopped before leaving the frame which makes it more robust in distinguishing foreground objects from the background. To estimate the road region



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

Fig. 1: Extracting the road region using the cumulative maps of the flood-fill method. (a) Original video frame. (b) The background obtained by the GFM method. (c) The edges of the background image. (d) The retrieved road map.

based on the moving vehicles we use both foreground and background information retrieved from the GFM method.

In traffic video surveillance applications that involve a stationary camera overlooking a highway, most if not all of the motion occurs on the road. Each time a vehicle passes along the road, the pixels of the road map in the corresponding location to its foreground mask are added by a constant positive value. In order to discard the faulty outputs of the foreground segmentation method, a tracking approach is utilized to only include the foreground mask of the moving vehicles and discard the pixels that are segmented as foreground due to the possible motion in the areas outside of the roadway. For the sake of simplicity and real-time performance we apply the blob-tracking method introduced in [6] for vehicle tracking. At each frame the foreground mask of each tracked vehicle is saved separately and if the life-time and moving length of that track exceeds predefined thresholds the corresponding pixels of the entire foreground mask of that track in the active traffic region map are added with a positive number. Applying filters to the foreground mask based on track life-time and moving length of each track ensures that only vehicles passing along the road are considered as part of the active traffic region and noises in the foreground mask are disregarded.

The accumulative foreground mask is used to construct a set of sampled road pixels represented by Ω_{sm}, and a probability map is generated based on color differences. A set of four-dimensional feature vectors is created that contains the blue, green, red, and hue values of each pixel:

$$\mathcal{F}^t = \{f_1^t, f_2^t, ..., f_N^t\} = \{f_i^t\}_{i=1}^N \tag{1}$$

where f_i^t is the D dimensional feature vector of pixel i at frame t. Then the standardized Euclidean distance between the mean value of pixels in Ω_{sm} and each feature vector are calculated as follows:

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

where $i = 1...N$ is the index of each pixel, t is the frame number, $D = 4$ is the number of features, $\overline{f_j^t}$ is the mean value of the j-th feature, f_{ij}^t is the j-th feature of pixel i, d_i^t is the standardized Euclidean distance, and $(\sigma_j^t)^2$ is the variance of the j-th feature. Then the probability map $\mathcal{P}_\mathcal{R}$ representing the probability of a pixel belonging to the road region is denoted by:

$$\mathcal{P}_\mathcal{R}{}^t = \{p_1^t, p_2^t, ..., p_N^t\} = \{p_i^t\}_{i=1}^N \tag{3}$$

where N is the total number of pixels, and $p_i \in [0, 1]$ is the probability value of pixel i at frame t which is in turn calculated by considering the empirical rule:

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

where $i = 1...N$ is the pixel index, $\overline{\sigma^t}$ is the mean standard deviation of the features among the D dimensions in Ω_{sm}, k is a natural number in $\{k \in \mathbb{N} | 1 < k \leq max(d_i^t - \overline{\sigma^t})\}$, and p_i is the resulting probability value of pixel i at frame t.

After calculating the probability value of pixels at each frame, the temporal fusing algorithm is applied in order to update the probability map in the following video frames as follows:

$$\begin{aligned} \hat{p_i}^t &= \frac{\sum_{f=1}^{t} w^f p_i^f}{1 + \sum_{f=1}^{t} w^f} \\ w^f &= |\Omega_{sm}^f| \end{aligned} \tag{5}$$

where $i = 1...N$ is the pixel index, w^f is the weight given to the frame f calculated based on the number of pixels in the aggregated sample mask, p_i^f is the probability value of pixel i at frame f, and $\hat{p_i}^t$ is the updated probability value of pixel i. A binary mask $\mathcal{P}_\mathcal{R}^*$ is obtained by applying the Otsu's threshold [9] on the resulting map.

The flood-fill algorithm considers the edges of the image, which is a reliable feature, especially in highway videos where the dominant road boundaries create strong edges. The Canny edge detection method is applied on the background image achieved from the GFM method 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 the grayscale background. Figure 1(c) represents the edges extracted from the background image. Utilizing the edge information, we apply the flood-fill algorithm with a connectivity value of 4 on the background image which fills the connected components with a given value in order to segment the road region. The maximal lower and upper intensity difference between the currently observed pixel and one of its four nearest neighbor pixels that belong to the same component, or a seed pixel that is being added to the component is calculated based on the standard deviation of the background image (see equation 6).

$$\begin{gathered} mean = \frac{1}{N}\sum_I src(I) \\ stddev = \sqrt{\frac{\sum_I (src(I) - mean)^2}{N}} \\ dif = max(1, \frac{stddev}{c}) \end{gathered} \tag{6}$$

where $mean$ is the mean value of the background image, N is the total number of pixels in the background image, src is the input array with 3 channels which refers to the background image obtained from the GFM method, I is the intensity value of each pixel, c is a pre-defined constant, and dif is the maximal lower or upper intensity difference. The maximal lower and upper thresholds are selected based on the general intensity difference among the pixels of the entire background image.

When the dissimilarity among intensity values is relatively large, the connected components in the Flood-Fill method tend to grow slower and thus a larger value for the maximal thresholds is chosen. On the other hand, in cases where the intensity values are close, e.g, foggy and rainy weather conditions or night time videos, the distinction level between pixels that belong to the road region and pixels that belong to the side of the road is lower. Therefore, in order to avoid connecting the pixels outside of the road area to the generated components, a smaller value is needed for the maximal thresholds. Another consideration for avoiding the inclusion of the pixels outside of the road area as seed points for flood-fill operation, a single seed point is selected for each vehicle based on its moving direction. We consider the moving direction of the vehicle and always select one of the corner points of its surrounding bounding box that is certain to belong to the road area and thus avoiding the selection of non-road pixels as seed points.

The temporal road map retrieved from the accumulative foreground mask of the moving vehicles is also utilized in order to avoid connecting pixels that are not involved in the active traffic zone to the components that represent the roadway. Since the road boundaries may not always be visibly distinguishable, the edges obtained from the background image are not always strong enough to stop the connection process of the seed-fill method. The flood-fill algorithm is limited to the zero pixels of a given binary mask and does not go across the non-zero pixels. Along with the edges of the background image, the inverse binary mask of the accumulative map of foreground pixels and the binary road mask $\mathcal{P}^*_{\mathcal{R}}$ is utilized in order to avoid including the non-road pixels if the filling has not already stopped at the edges. Therefore, the output of the flood-fill method is refined by excluding the regions that involve low to no motion and the regions with different features. Figure 1 presents examples of the flood-fill algorithm applied on traffic videos in a period of one minute.

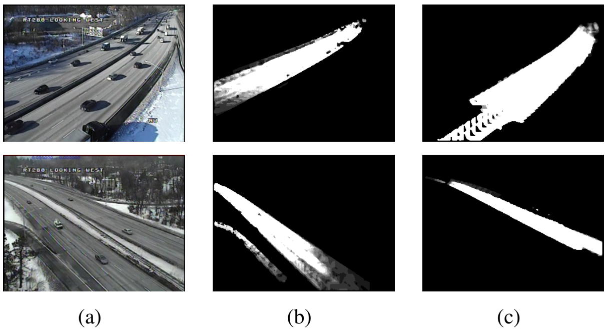


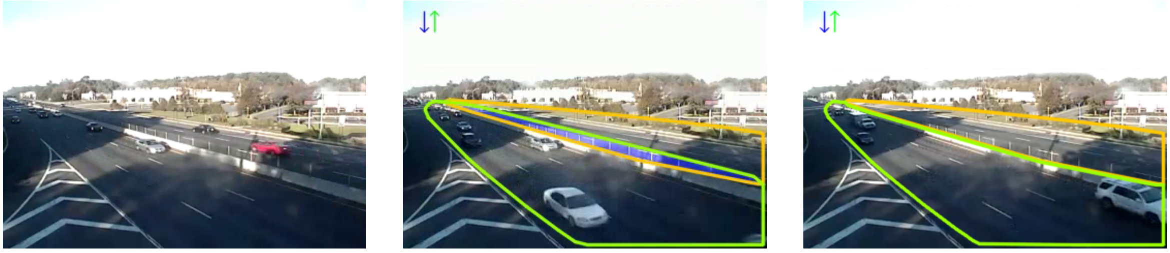
Fig. 2: Separated accumulative foreground masks of the moving vehicles. (a) Original traffic video frame. (b) and (c) Accumulative foreground masks the left and right sides, respectively.

B. A novel statistical method for separating major traffic directions

Most roads and highways carry traffic in two major opposite directions. In case of most traffic video analytic tasks a separate RoI is needed for each side of the road. In order to retrieve an RoI for each side of the road the tracking information obtained from the blob-tracking approach is used to detect the moving direction of each vehicle. The centroid of each track at the starting and ending position is compared to estimate the direction of its movement. To avoid the effects of noises in the foreground and noisy results of the tracking method, only vehicles with high enough movement size and speed are considered. Each time such a vehicle passes along the road, the pixels with a corresponding location to its foreground mask are added with a positive number in the road map of the correct direction and added with a negative number in the road map of the opposite direction. To avoid having common areas between left and right regions we try to remove the foreground mask of a vehicle from the opposite side when it is being added to one side in case it has previously been added to the opposite side by mistake.

For each tracked vehicle that passes along the road, the left and right sides of the road are updated as follows:

$$m = max(0, \alpha \sum_{f=1}^{T} m_v - \beta(m_o + \sum_{f=1}^{T} m_{vo})) \tag{7}$$



(a) (b) (c)

Fig. 3: Assigning the overlapping area between the maps of the two traffic direction to the correct side. (a) The original traffic video frame. (b) The blue color indicates the overlapping area between two RoIs. (c) The overlapping area is assigned to the correct RoI and removed from the other RoI.

where m is the traffic region map for one side of the road, m_v is the foreground mask of the vehicle passing along that side at frame f, m_{vo} is the foreground mask of the vehicle passing along the opposite side at frame f, m_o is the traffic region map of the opposite side of the road, T is the current frame, and α and β are predefined coefficients between 0 and 1. In order to speed up the update process of the traffic region maps, α and β should be closer to 1 and in order to reduce the update errors they should be closer to 0. Each road map is then updated by applying Otsu's threshold:

$$m_f = \begin{cases} 1, & \text{if } m_{acc} \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where m_f is the final traffic region binary map for each side, m_acc is the accumulative foreground masks in that side, f is the current frame, τ is the calculated Otsu's threshold, m_f is the foreground mask of frame f, and F is the total number of frames. The Otsu's threshold is applied to remove noises that are mostly caused by occasional noises in the foreground mask. Figure 2 shows examples of the separated accumulative foreground masks for the two major directions of the traffic flow.

In order to obtain an RoI for each side of the road that contains the road itself and a good portion of its surroundings, the convex hull of the road map's contour is used for each side. The two convex hulls corresponding to the contours obtained from the road map of each side of the road have proven to be good representations for the RoI, as they involve the entire road and its surroundings while avoiding the regions outside of the road and therefore saving the video analytic applications from the unnecessary noises and computational overload. However, in videos where the camera angle is from one side of the road, the foreground mask of the vehicles from different sides can overlap each other which in turn causes an intersecting area between the convex hulls of the two sides in the middle part of the road. The overlapping area should be removed from the RoI of the wrong side to avoid false positive results in further video analysis tasks. In order to decide which side of the road the overlapping area belongs to, the intersection between the overlapping area and the convex hull of each side is calculated and the overlapping area is removed from the RoI of the side with lower intersection. Figure. 3 shows the overlapping area removed by our proposed method.

In some videos the traffic flows in more than two directions and further steps are required to be taken in order to extract only the regions corresponding to the major directions and exclude others. In this case, using the direction obtained from tracking is not enough to separate the regions with similar direction but of different road segments. Here, we have applied a statistical method based on Gaussian Mixture Models (GMM) in order to estimate the general moving velocity of the vehicles at various locations of the road. At each frame, the Lucas-Kanade optical flow method [2] is applied to obtain a matrix of flow vectors in the size of the entire frame. The Lucas-Kanade optical flow method has incorrect outputs, specially in video with low resolution and the results of a few frames are not reliable for estimating the motion vectors. To overcome this problem, the non-zero magnitude and speed of the optical flow vectors in a sequence of frames are utilized as two-dimensional input vectors by the GMM method in order to estimate the most probable velocity at each pixel. The Gaussian modeling of the optical flow vectors is described as follows:

$$P(\mathbf{x}) = \sum_{k=1}^{K} W_k N(\mathbf{x}|\omega_k) \tag{9}$$

$$N(\mathbf{x}|\omega_k) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_k)^t \Sigma_k^{-1} (\mathbf{x}-\mu_k)\right\}}{(2\pi)^{d/2} \mid \Sigma_k \mid^{1/2}} \tag{10}$$

$$\sum_{k=1}^{K} W_k = 1 \tag{11}$$

where $\mathbf{x} \in \mathbb{R}^d$ is the two-dimensional feature vector containing flow angle and magnitude of each pixel, K is the number of Gaussian distributions in the flow model, W_k is the weight of the k_{th} Gaussian distribution $N(\mathbf{x}|\omega_k)$. μ_k and Σ_k are the mean vector and the covariance matrix of the k_{th} Gaussian density $N(\mathbf{x}|\omega_k)$. Note that the Gaussian model of each pixel is updated only when the magnitude of the optical flow is larger zero. The results of the GMM is further refined by

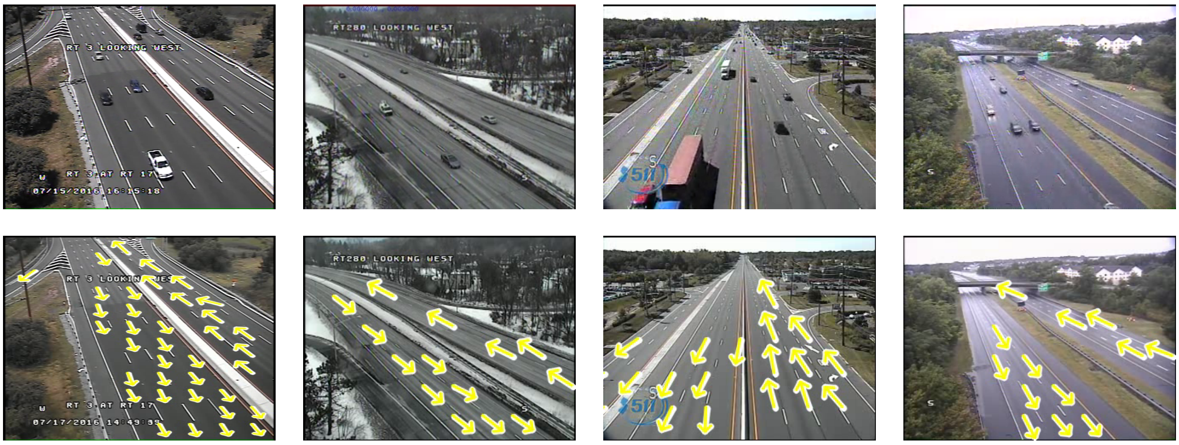
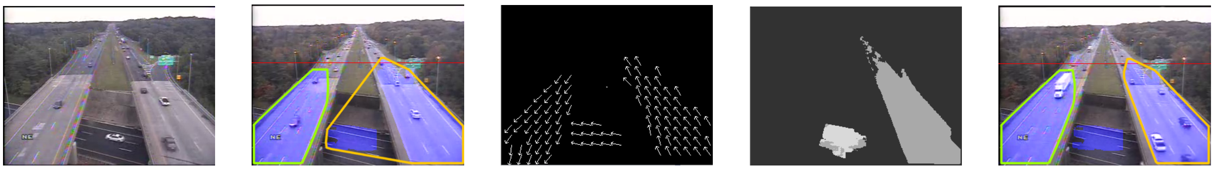


Fig. 4: Extracting a matrix of motion flow vectors using GMM method with optical flow vectors as input. First row contains sample frames of traffic videos. Second row represents the corresponding flow model matrix obtained from the GMM method.



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

Fig. 5: Excluding smaller road regions with similar direction to one of the major traffic regions. (a) The original traffic video frame. (b) The road under the bridge is incorrectly grouped with one of the major traffic regions. (c) The flow vectors obtained by the GMM method. (d) Applying Kmeans clustering method to separate the small region with a similar direction. (e) The small region is excluded from the RoI.

removing incorrect estimations based on the general direction of each tracked vehicle. Figure 4 shows examples of the optical flow vectors modeled by the GMM method.

After generating the flow matrix, the K-means clustering approach is applied in order to group pixels with vectors of close angles together and thus excluding the regions that are not part of the two major traffic directions. Figure 5 shows an example of how the smaller region falsely included in one of the two major traffic regions is separated and removed from the RoI.

IV. Experiments

In this section, we introduce the dataset used and present some experimental results to evaluate the performance of our proposed novel Region-of-Interest determination method. We further discuss the strengths and limitations of the RoI determination approach.

A. Dataset

As the automatic two direction RoI determination method is a relatively new topic in traffic video processing, there is no publicly available benchmark dataset with ground-truth data for two-side roadways. We have used real traffic video sequences from the New Jersey Department of Transportation (NJDOT) for evaluation. This dataset contains dozens of diverse traffic surveillance video scenarios, with different illumination circumstances, weather conditions, and spatial resolutions.

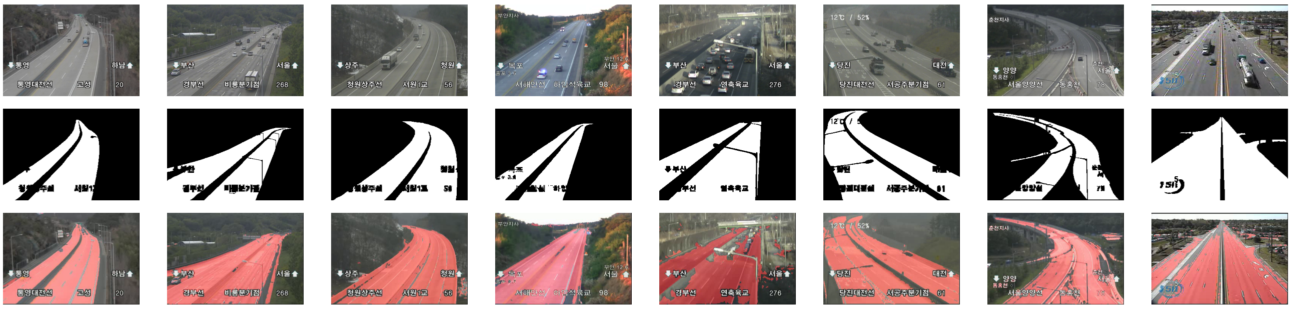
B. Performance analysis

The hardware used to implement the method is a desktop with an Intel Core i7-8700 Processor. The running speed of our proposed method is 38.7 frames per second (fps) which is fast enough to be used as a pre-processing step in real-time traffic video analysis tasks.

The quantitative performance of the entire road region recognition introduced in III-A is evaluated by the following metrics:

$$\begin{cases} FPR = F_P/(F_P + T_N) \\ PRE = T_P/(T_P + F_P) \\ REC = T_P/(T_P + F_N) \\ F_1 = 2 \times (PRE \times REC)/(PRE + REC) \end{cases} \tag{12}$$

where T_P, F_P, T_N and F_N are the number of true positive, false positive, true negative, and false negative classified pixels, respectively. FPR, PRE, REC, and F_1 refer to false positive rate, precision, recall, and F1-score, respectively. Table I illustrates the quantitative performance of the road region recognition method for the 8 sample traffic videos.

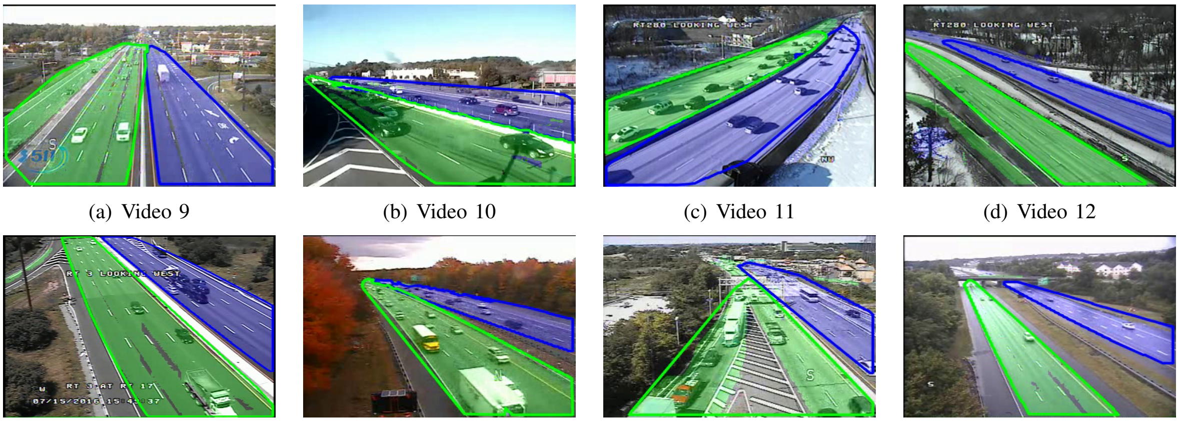


(a) Video 1 (b) Video 2 (c) Video 3 (d) Video 4 (e) Video 5 (f) Video 6 (g) Video 7 (h) Video 8

Fig. 6: Road extraction results in traffic videos with regular and 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.

TABLE I: The quantitative evaluation of the proposed road recognition method

Video #	1	2	3	4	5	6	7	8	**Average**
Precision	0.98	0.87	0.88	0.92	0.89	0.97	0.80	0.99	0.91
Recall	0.96	0.93	0.94	0.93	0.92	0.89	0.91	0.73	0.90
F-Score	0.96	0.90	0.94	0.93	0.92	0.93	0.86	0.84	0.91



(a) Video 9 (b) Video 10 (c) Video 11 (d) Video 12

(e) Video 13 (f) Video 14 (g) Video 15 (h) Video 16

Fig. 7: Some experimental results of the proposed method on traffic surveillance videos. The blue color and green color indicate the two sides of traffic (RoIs) determined by our proposed method.

Figure 6 represents a few samples of the extracted road regions. Small under-segmentation and leak segmentation errors can be seen in Figures 6(e) and 6(h) and Figures 6(b), 6(d) and 6(g), respectively which can be overlooked in terms of having insignificant negative effects on the determined RoI.

The performance of the RoI detection method introduced in III-B is evaluated using videos with various view angles and illumination conditions. Table.II shows the video information we have used in our experiment. Figure 7 illustrates some examples of the RoIs determined by our proposed method. In each frame, the green and blue colors represent the two traffic regions (RoI) determined by our proposed method respectively. We can see that the automatically detected regions covers most of the road regions, which can directly be utilized as the RoIs in the applications of traffic surveillance videos.

C. Discussion

In this paper, we have not made any assumptions about the shape prior of the roadway or the viewing angle of the camera. This approach can work on straight, curved, fork, and other road structures. Also, the method is completely automatic and performed in real-time which makes it applicable in realworld scenarios. However, in some videos with challenging illumination and weather conditions, the initial road region extraction might have leak segmentation errors due to the similarities between the road pixels and the surrounding (e.g. sky). These errors are later dealt with by using the location of moving foreground objects, however achieving a good RoI determination can take longer.

TABLE II: The properties of the traffic video sequences represented in the figure 7

Video #	9	10	11	12	13	14	15	16
Resolution	320 × 240	352 × 240	640 × 482	640 × 480	352 × 240	320 × 240	320 × 240	320 × 240
FPS	15	15	15	15	30	15	15	15

V. Conclusion

Determining the region of interest (RoI) is a fundamental pre-processing step in video analysis applications. In this paper, a statistical method is proposed to automatically determine the region of interest corresponding to two major traffic directions in surveillance videos captured from roads with bidirectional traffic. Our proposed method has two contributions. First, the road region is segmented automatically by using color, edge, and temporal features and applying a background subtraction method along with the flood-fill operation. Second, two regions of interest are generated representing the major traffic directions in roads and highways with bidirectional traffic. As opposed to the supervised learning methods, the proposed method can adapt well to a wide range of videos with illuminations conditions and viewing angles in real-time. The experimental results using real traffic videos provided by NJDOT demonstrate good performance of the proposed methods.

Acknowledgment

This paper is partially supported by the NSF grant 1647170.

References

[1] E. J. Almazan, Y. Qian, and J. H. Elder, "Road segmentation for classification of road weather conditions," in *European Conference on Computer Vision*. Springer, 2016, pp. 96–108.

[2] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

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

[4] P. Chandran, M. John, N. Mithilesh *et al.*, "Road tracking using particle filters for advanced driver assistance systems," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 1408–1414.

[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] G. Cheng, Y. Qian, and J. H. Elder, "Fusing geometry and appearance for road segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017.

[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] 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.

[10] 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.

[11] M. Junaid, M. Ghafoor, A. Hassan, S. Khalid, S. A. Tariq, G. Ahmed, and T. Zia, "Multi-feature view-based shallow convolutional neural network for road segmentation," *IEEE Access*, vol. 8, pp. 36 612–36 623, 2020.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] H. Li, Y. Chen, Q. Zhang, and D. Zhao, "Bifnet: Bidirectional fusion network for road segmentation," *arXiv preprint arXiv:2004.08582*, 2020.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] ——, "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.

[23] W.-S. Shin, D.-H. Song, and C.-H. Lee, "Vehicle classification by road lane detection and model fitting using a surveillance camera." *JIPS*, vol. 2, no. 1, pp. 52–57, 2006.

[24] 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.

[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] 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.

[27] 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.

[28] 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.

[29] 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.