A Novel Global Foreground Modeling Guided Background Modeling Method for Real Time Foreground Detection in Video

Anonymous WACV submission

Paper ID 1019

Abstract

A novel global foreground modeling (GFM) guided background modeling method, which is capable of detecting stopped foreground objects, is presented in this paper. Specifically, the GFM method, which detects both the moving and stopped foreground objects, guides the background modeling method, which detects only the moving objects, for adaptive background updating. The contributions of the paper are three-fold. First, a novel GFM guided background modeling method is proposed by adaptively updating the background based on the difference between the two foreground masks produced by the GFM method and the background modeling method, such as the Zivkovic's method, respectively. As a result, the proposed method both improves the foreground detection and detects the stopped foreground objects. Second, a boosting strategy is integrated to the proposed method for eliminating the false alarms caused by noise. Third, real traffic videos are used to evaluate the effectiveness of the proposed method. In particular, experimental results using the real traffic videos from the New Jersey Department of Transportation (NJ-DOT) show that the proposed GFM guided background modeling method is able to detect stopped foreground objects, such as stopped vehicles, in real time.

1. Introduction

In video analysis, foreground object detection is one of the most extensively studied topics. Many methods have been proposed to detect the foreground objects [1], [2], [3], [4], [5]. One of the challenging issues in foreground object detection is the detection of the stopped moving foreground objetcs. Most foreground detection methods cannot keep detecting the foreground objects when they stop moving, or can only detect those objects for a short period of time. However, some real world applications are interested in those stopped moving foreground objects, and need the foreground detection methods to detect them, such as the industrial production line monitoring system, and the traffic incident detection system.

In this paper, we propose a novel GFM guided background modeling method, which can improve the foreground detection and detect the stopped foreground objects in real time. In addition, we apply our proposed method to detect the stopped moving vehicles in the real traffic surveillance videos. The contributions of this paper are three-fold.

First, we present an adaptive background updating strategy based on the global foreground modeling (GFM) method [6], [5]. The GFM method has the capability to detect the temporarily stopped foreground objects, while the Zivkovic's method [7], [8] does not. However, the stopped foreground object detected by the GFM method will eventually be detected as background with the updating of the background model. In order to detect the stopped foreground objects more accurate and robust, we use an adaptive background updating strategy to update the background model. We use the difference between the two foreground masks produced by the GFM method and the Zivkovic's method as an updating matrix, and adaptively update the background model for those stopped foreground object areas. As a result, the proposed method both improves the foreground detection and detects the stopped foreground objects.

Second, a boosting strategy is integrated into the proposed method to reduce the false alarms when detecting the stopped foreground objects. Heuristics or criteria from prior knowledge are used to define some weak classifiers. There are several kinds of heuristics that can be used regarding the motion, size, edge, and color, etc. As any individual heuristic is not sufficient for differentiating the target objects and the noise, we utilize several heuristics by means of integrating a boosting strategy, such as the Adaboost method [9], to enhance the classification performance.

Third, we apply the proposed method on the real videos from the New Jersey Department of Transportation (NJ-DOT) to illustrate the effectiveness of our method. The experimental results demonstrate the advantage of the proposed GFM guided background modeling method in detecting stopped foreground objects, such as stopped vehicles, in real time.

2. Related Work

Many statistical modeling methods have been proposed for foreground object detection [1], [2], [3], [4]. In 1999, Stauffer and Grimson proposed a Gaussian Mixture Model (GMM) method to estimate the background model [10], and further detect the foreground pixels. Based on their research, Hayman and Eklundh used several signification Gaussian densities in the GMM as the background model, and use the residual as the foreground model to subtract the background from the videos [11]. Later, Zivkovic presented another statistical modeling method that uses the GMM method to model the background and uses a uniform distribution to model the foreground [7], which is integrated in opencv and is widely used in industry. More recent, Hofmann et al. proposed a pixel-based adaptive segmentation (PBAS) method [12]. By using adaptive threshold for each pixel, the PBAS method can better adapt to illumination changes.

Some region based foreground detection methods are also proposed to detect the foreground objects. The intention of this kind of methods is to improve the foreground detection performance by increasing the dimensionality of the feature vector. Wren et al. presented a blob-based method to detect and track human body [13]. Pandey and Lazebnik used a deformable part-based model for feature extraction and trained a support vector machine (SVM) for foreground classification [14]. Varadarajan et al. proposed a region based method [15], which uses small blocks as unit to extract feature and detect the foreground. Qin et al. used a basic matrix background modeling method that estimates the basis matrices of the background [16].

Note that these methods have an issue that they are not able to detect the stopped foreground objects. Facing this problem, some improved methods are proposed. Zhong et al. propsoed an adaptive background modeling method based on the PBAS method [17]. By using a counter to control the updating speed of the background model, some stopped foreground objects can be detected for a period of time. Shi and Liu proposed a Global Foreground Modeling (GFM) method, which is able to detect the temporarily stopped foreground objects as well [6], [5].

Some foreground detection methods based on deep neural network are also proposed, such as the convolutional neural networks (CNN). These methods use some supervised learning models to address the foreground detection problem [18], [19], [20], [21]. This kind of mehtods detect the foreground objects based on automatically learnt features, and are able to detect the foreground objects no matter they are moving or not. However, these methods need huge amount of labeled training data to train the parameters, while the labeled data is hard to achieve. There are still some other challenging issues with these deep learning methods, such as the generalization performance, running speed, etc.

Our proposed GFM guided background modeling method, in contrast to the above methods, is capable of detecting both the moving and stopped foreground objects in real time.

3. A Novel GFM Guided Background Modeling Method

In this section, we first briefly review the GFM foreground detection method [6], [5] and the Zivkovis's foreground detection method [7], [8]. We then present the novel GFM guided background modeling method, which is capable of detecting stopped foreground objects.

In the GFM method, the background and the foreground models are built separately, and the Bayes decision rule for minimum error is used to classify each pixel in a frame into the foreground class or the background class. For background modeling, a traditional Gaussian Mixture Model (GMM) $p(\mathbf{x}|\omega_{gmm,i,j})$ is built for every location in a frame [10]. For notational simplicity and without loss of generality, we will drop the subscripts i, j in the following equations. At time t, the probability density function at location (i, j) is estimated as follows:

$$p_t(\mathbf{x}|\omega_{gmm}) = \sum_{k=1}^{K} \alpha_{k,t} N\left(\mathbf{M}_{k,t}, \Sigma_{k,t}\right) \tag{1}$$

where $\mathbf{x}$ is the pixel value, ω_{gmm} means the Gaussian mixture model, K indicates the number of Gaussian densities in the GMM method, $\alpha_{k,t}$ is the weight, and $N\left(\mathbf{M}_{k,t}, \Sigma_{k,t}\right)$ presents a Gaussian distribution with the mean vector $\mathbf{M}_{k,t}$ and the covariance matrix $\Sigma_{k,t}$. The K Gaussian density functions are sorted in a descending order according to the weights $\alpha_{1,t}, \alpha_{2,t}, \cdots, \alpha_{K,t}$.

The most significant single Gaussian density $N\left(\mathbf{M}_{1,t}, \Sigma_{1,t}\right)$, which has the largest weight, is choosen as the conditional probability density function $p_t(\mathbf{x}|\omega_b)$ for background, where ω_b stands for the background. And also the mean vector $\mathbf{M}_{1,t}$ is used to estimate the background value for time instant t. In every frame, the pixel value at location (i, j) is used to update the Gaussian mixture model.

For foreground modeling, a global statistical model is built. L Gaussian density functions $p(\mathbf{x}|\omega_1)$, $p(\mathbf{x}|\omega_2)$, $\cdots$, $p(\mathbf{x}|\omega_L)$ are estimated as conditional probability density functions for all the foreground pixels in the global foreground model, respectively. Further more, the conditional probability density function for the foreground is chosen by using the Bayes classifier:

$$\omega_f = \arg\max_{\omega_i} \{p(\mathbf{x}|\omega_i) P(\omega_i)\} \tag{2}$$

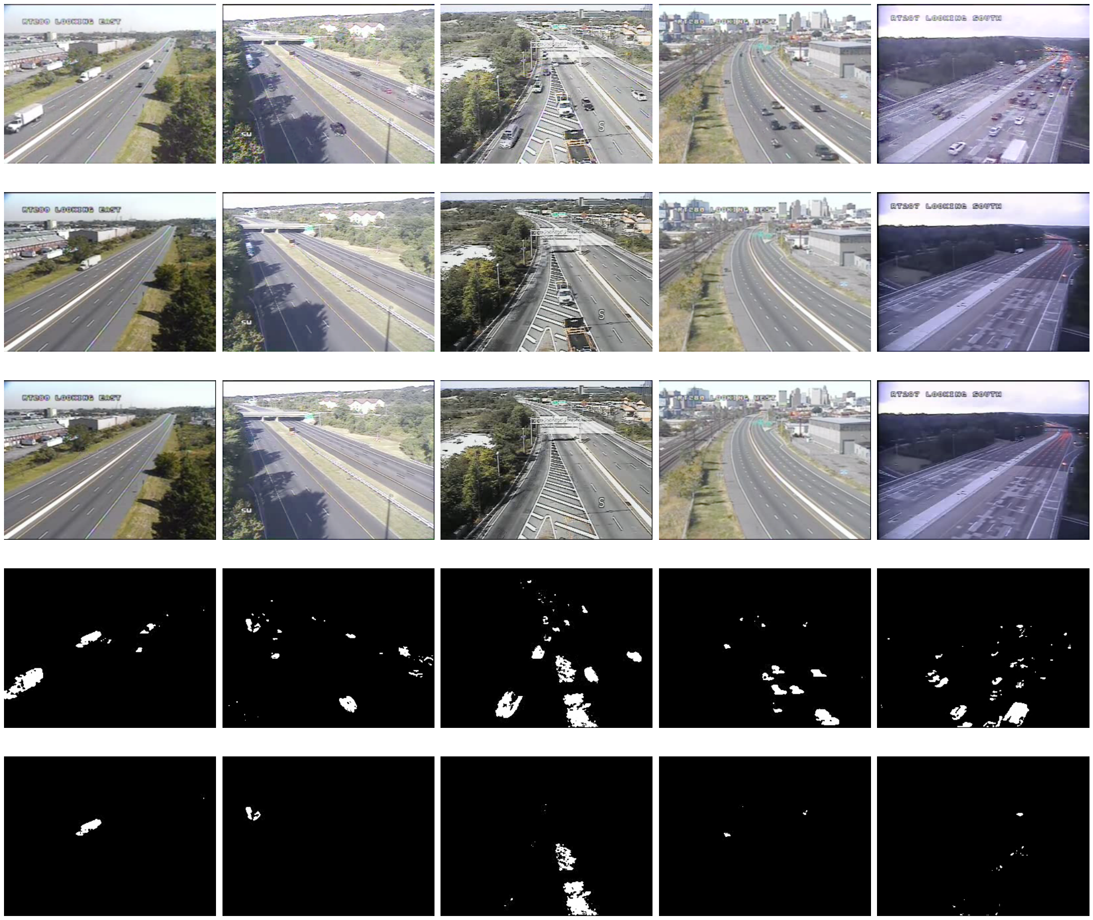


Figure 1. The foreground detection and background estimation performance of the proposed method. The first row shows some video frames from the NJDOT traffic videos with a spacial resolution of 352×240. The second row shows the background estimated using the GMM method. The third row shows the background estimated using our novel GFM guided background modeling method. The fourth row shows the foreground mask detected by our method. The fifth row shows the stopped foreground objects detected by our method.

where ω_f presents the foreground class, and $\omega_i \in \omega_1, \omega_2, \ldots, \omega_L$.

Finally, the Bayes decision rule for minimum error [22] is applied to classify each pixel into the foreground class or the background class using the following discriminant function:

$$c(\mathbf{x_t}) = p_t(\mathbf{x_t}|\omega_f)P_t(\omega_f) - p_t(\mathbf{x_t}|\omega_b)P_t(\omega_b) \tag{3}$$

where $p_t(\mathbf{x}|\omega_f)$, and $p_t(\mathbf{x}|\omega_f)$ are the estimated conditional probability density functions for the foreground and background at time t, $P_t(\omega_f)$ and $P_t(\omega_b)$ are the prior probabilities of the foreground and background, respectively. The pixel $\mathbf{x_t}$ is classified to the foreground class if $c(\mathbf{x_t}) > 0$, and to the background class otherwise.

After the classification, we assign all the foreground locations to 1, and all the background locations to 0. Thus we can get a binary foreground mask. This foreground mask contains both the moving foreground objects and the stopped moving foreground objects, we call it the strong foreground mask F_s.

In the Zivkovic's method, the probability density function of each pixel at location (i, j) is also estimated using the Gaussian Mixture Model (GMM) [10], and the first M components are used as background density function $p(\mathbf{x}|\omega_b)$.

At time t, if pixel $\mathbf{x}_t$ satisfies:

$$p(\mathbf{x}_t|\omega_b) > C_{thr} \tag{4}$$

where C_{thr} is a threshold value, it will be classified as a background pixel, otherwise it is classified as a foreground pixel.

After that, we assign all the foreground pixels to 1, and background pixels to 0 and get a binary foreground mask. This mask only contains the moving foreground objects, we call it the weak foreground mask F_w.

Even though the GFM method is able to detect those stopped foreground objects, but as time goes on, the es-

Table 1. The description of the video sequences we used in our experiments

	video resolution	frame rate (fps)	length of the video (mins)	bit rate (kbps)	special condition	number of videos
1	320×240	15	30	45 to 132	normal	12
2	320×240	15	30	41 to 131	night time	4
3	320×240	15	30	47 to 176	strong shadow	9
4	320×240	15	30	123	fog	1
5	320×240	15	30	82	rain	1
6	640×480	15	30	1066	snow	1
7	352×480	30	60	633 - 839	normal	2
Total						30

timated background will be affected by the stopped foreground objects. The foreground object will become a fake background at their stopped position. Therefore, the stopped objects will disappear in the foreground mask eventually. To solve this problem, we propose a novel background modeling method, which will improve the stopped foreground objects detection performance of the GFM method.

As we mentioned above, the strong foreground mask F_s includes both the moving and stopped foreground pixels, while the weak foreground mask F_w includes only the moving foreground pixels. We use the difference between these two masks to locate the stopped foreground pixels. If a pixel is detected as foreground in the F_s, but as background in the F_w, we regard that pixel as a stopped foreground pixel. At time instant t, the value of the candidate stopped foreground mask at location (i, j) is defined as follows:

$$F_{stop}(t) = F_s(t) > F_w(t) ? 1 : 0 \tag{5}$$

where F_{stop} indicates the stopped foreground mask, 1 means the location is a stopped foreground object, and 0 means the other foreground objects or background. For specific tasks, we will further filter this stopped foreground mask to remove some noises using the boosting strategy we will introduce in Sec. 4.

The stopped foreground mask F_{stop} contains the stopped foreground objects that we are interested in, so we do not want them be considered as background. Therefore, we use an adaptive background updating strategy to update the Gaussian mixture background model. The adaptive updating strategy is as follows:

$$p_t(\mathbf{x}|\omega_{gmm}) = \begin{cases} \sum_{k=1}^{K} \alpha_{k,t} N\left(\mathbf{M}_{k,t}, \Sigma_{k,t}\right), F_{stop}(t) = 0 \\ p_{t-1}(\mathbf{x}|\omega_{gmm}), F_{stop}(t) = 1 \end{cases} \tag{6}$$

where $\alpha_{k,t}$, and $N\left(\mathbf{M}_{k,t}, \Sigma_{k,t}\right)$ are updated following the rule described in [10]. By using this novel background updating strategy, the stopped foreground objects can be detected as long as they stop there.

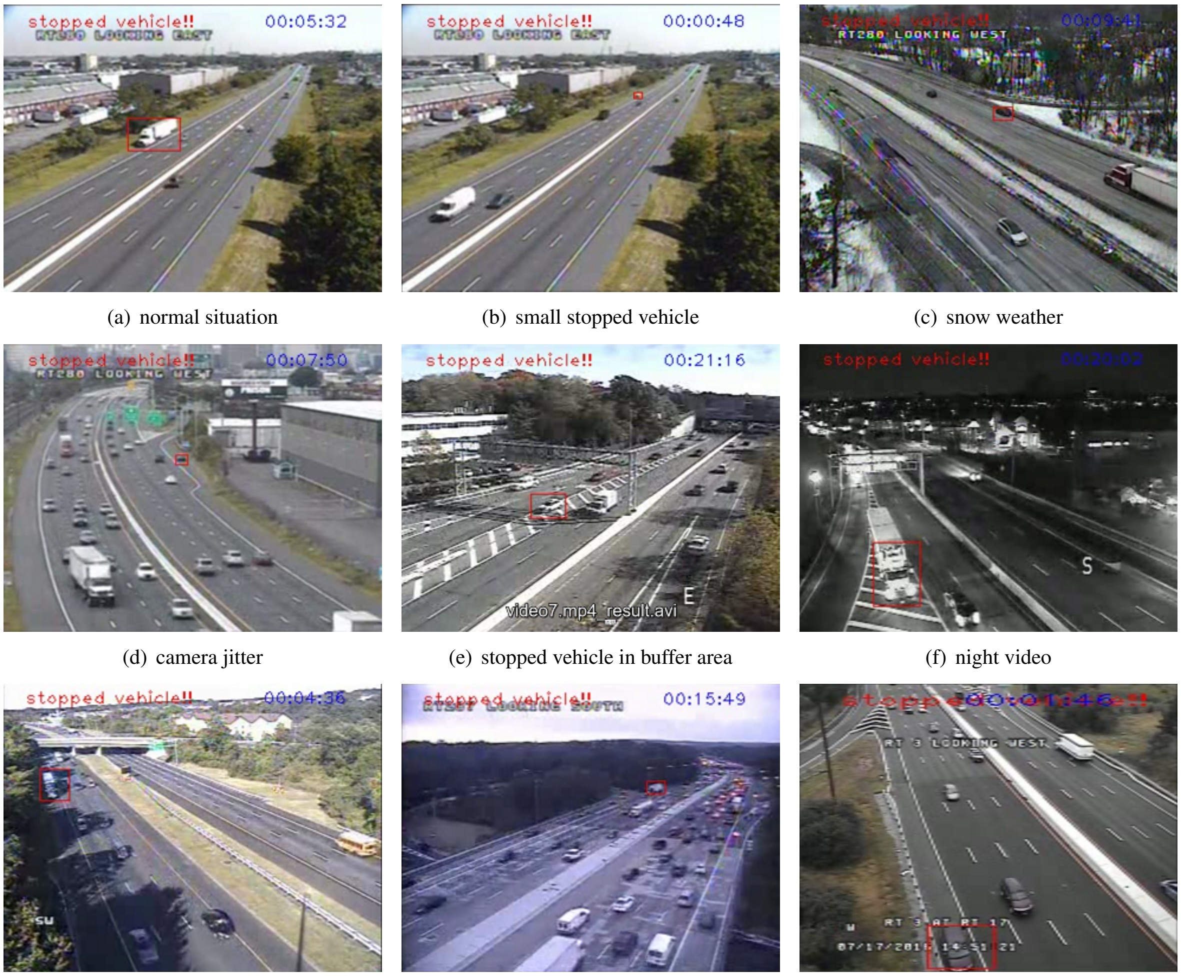
Figure 1 shows some foreground masks and the corresponding background estimated by different method. The first row shows some video frames from the NJDOT traffic videos with a spacial resolution of 352×240. The second row shows the background estimated using the traditional Gaussian mixture model. The third row shows the background estimated using our novel GFM guided background modeling method. The fourth row shows the foreground mask detected by our method. The fifth row shows the stopped foreground object mask detected by our method. We can see the stopped foreground objects are merged into the background in the GMM method, but our proposed GFM guided background modeling method can get a clean background.

4. Boosting the Performance of Foreground Detection and False Alarms Reduction

Due to many real world conditions, such as the shaking tree leaves, the shadow of the utility poles, the reflection of lights, and the noise due to camera jitters or bad weather, the stopped foreground detection is not 100% accurate. Besides the target stopped foreground objects, some other objects are detected as stopped foreground objects, which cause false alarms. To eliminate these false alarms, we propose a boosting strategy based on a number of criteria for more accurate stopped foreground objects detection. The criteria are defined by geometric and statistical features, such as the normal size of the objects, the mean and variance values, and the edge information. These criteria are integrated to boost the performance of foreground detection and false alarms reduction by means of the boosting strategy, such as the Adaboost method [9].

In particular, we first assign each connected area in the candidate stopped foreground mask to one block, and use the morphological operations to connect the closed blocks. We then introduce some weak classifiers based on some criteria. We finally use a boosting strategy, such as the Adaboost method [9], to classify each block into a target object class or non-target class.

The first criterion, which is based on geometric features, is defined by the target normal size. Specifically, the size of the target object should be within a reasonable range. For example, the size of a vehicle can be various, but it is still



(a) normal situation (b) small stopped vehicle (c) snow weather

(d) camera jitter (e) stopped vehicle in buffer area (f) night video

(g) strong shadow (h) illumination change (i) stopped vehicle behind the caption

Figure 2. Some stopped vehicle detection results. The stopped vehicles detected by our proposed method are marked with the red rectangles.

in a limited range. A vehicle can never be smaller than a pedestrian or a motorcycle, or several times bigger than a big truck. To eliminate the influence of the perspective view to the size of the objects, we warp the frame to normalize the size of each object into the same scale, and use the average size of the target objects to filter out some blocks that are too small or too big.

The second criterion is define by statistical features. Note that the mean and variance values of the object area are always different from those of the background region. Due to the camera jitter or illumination changes, some background areas sometimes are classified as stopped foreground objects, but this kind of areas always has different statistical features (e.g. mean and variance values) from the target objects. We calculate the variance and the mean values of the original frame and the estimated background at each block location, and the variance and the mean values of the difference between the original frame and the estimated background. By using the difference between the variance and the mean values of the original frame and the estimated background, and the variance and the mean values of their difference, the shadows or some noise can be differentiated from the target objects.

The third criterion is define by geometric features, namely, edges. Note that the edge information of the foreground object and the false alarm region is often different. We use canny edge detection method [23] to extract the edge information from the frame and the estimated background. Then we calculate the similarity between the edges of the frame and the edges of the estimated background to determine if there is a target stopped object in the area, or just a noise. In addition, we use the edges exist in the frame but not in the background as the edges of the foreground object. By considering the density of the edge pixels, we can determine if the stopped object is our target.

These criteria may be used as weak classifiers for stopped foreground object detection, as usually the discriminatory power of these individual classifiers is weak. We therefore integrate these weak classifiers to create a stronger classifier by using a boosting strategy, such as the Adaboost method [9], for boosting the performance of foreground detection and false alarms reduction.

5. Experiments

We show experimental results and analysis the performance of our proposed method in this section. Specifically, we apply our proposed GFM guided background modeling method on a stopped vehicle detection task in traffic surveillance videos to illustrate the usage and performance. The video sequences we tested are from the New Jersey Department of Transportation (NJDOT). These video sequences contain different real traffic situations and video qualities, such as low video resolution, bad video quality, camera jitter, and bad weather conditions. Most of these real world videos does not have good quality, this cause most of the vehicle detection methods can not identify the vehicles. The detail description of the video sequences are shown in Table. 1.

Among these videos, the stopped vehicle incident occurred 22 times, ranging from 10 seconds to 15 mins. The long stopping time causes some other foreground detection methods cannot keep detecting the stopped vehicles, even they can detect some temporarily stopped foreground objects. On contrast, our proposed method is able to detect those stopped vehicles as long as they are stopping there. Our method is able to detect 21 out of 22 of these stopped vehicle incidents, and no false positive detection occurred. The only one we miss is because of the night vision and highly blurred video frames.

The computer we use is a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. We implemented our proposed method using opencv in C++. In practice, we resize the videos with the resolutions of 640×480 and 352×480 to 320×240 and 176×240, respectively. The running times of our stopped vehicle detection method are shown in Table. 2. It shows that our method is able to process the videos in real time.

Table 2. The processing speed of our proposed method

	video resolution	processing speed (fps)
1	320×240	49
2	176×240	65

We show some stopped vehicle detection results in Fig. 2. These figures include several challenging conditions in real world traffic videos, such as low resolution, bad weather condition, camera jitter, night video, shadow, etc. We can see that our proposed method can detect the stopped vehicles under all these conditions without false positive detections.

6. Conclusion

We have presented in this paper a novel global foreground modeling (GFM) guided background modeling method for foreground detection. The contributions of this paper are summarized below. First, a novel GFM guided background modeling method is proposed, which is able to detect the stopped foreground objects. By adaptively updating the background model, the estimated background is more clean. In addition, the long time stopped foreground objects can be better detected. Second,we propose a boosting strategy to further enhance the stopped foreground detection. Some heuristics are used to build the weak classifiers. By boosting the discrimination power of these weak classifiers, the noises can be eliminated from the foreground mask. Third, we apply our proposed method to solve a real world problem. By taking advantage of our proposed method, that is able to detect the stopped foreground object stably and reliably, we detect the stopped vehicle incidents in the real world traffic surveillance videos. In particular, the real world traffic videos from the New Jersey Department of Transportation (NJDOT) are used in our experiments. The experimental results show that the proposed GFM guided background modeling method is able to detect stopped foreground objects, such as stopped vehicles, in real time.

References

[1] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.

[2] S. Brutzer, B. Hferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR 2011*, June 2011, pp. 1937–1944.

[3] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[4] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Computer Science Review*, vol. 23, pp. 1–71, 2017.

[5] H. Shi and C. Liu, "A new global foreground modeling and local background modeling method for video analysis," in *International Conference on Machine*

Learning and Data Mining in Pattern Recognition. Springer, 2018, pp. 49–63.

[6] ——, "A new foreground segmentation method for video analysis in different color spaces," in *24th International Conference on Pattern Recognition*. IEEE, 2018.

[7] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004*. IEEE Computer Society, 2004, pp. 28–31.

[8] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[9] Y. Freund, R. E. Schapire *et al.*, "Experiments with a new boosting algorithm," in *icml*, vol. 96. Citeseer, 1996, pp. 148–156.

[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *1999 Conference on Computer Vision and Pattern Recognition (CVPR '99), 23-25 June 1999, Ft. Collins, CO, USA*. IEEE Computer Society, 1999, pp. 2246–2252.

[11] E. Hayman and J. Eklundh, "Statistical background subtraction for a mobile observer," in *9th IEEE International Conference on Computer Vision (ICCV 2003), 14-17 October 2003, Nice, France*. IEEE Computer Society, 2003, pp. 67–74.

[12] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2012, pp. 38–43.

[13] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.

[14] M. Pandey and S. Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," in *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. V. Gool, Eds. IEEE Computer Society, 2011, pp. 1307–1314.

[15] S. Varadarajan, P. C. Miller, and H. Zhou, "Region-based mixture of gaussians modelling for foreground detection in dynamic scenes," *Pattern Recognition*, vol. 48, no. 11, pp. 3488–3503, 2015.

[16] M. Qin, Y. Lu, H. Di, and W. Huang, "A background basis selection-based foreground detection method," *IEEE Transactions on Multimedia*, vol. 18, no. 7, pp. 1283–1296, July 2016.

[17] Z. Zhong, B. Zhang, G. Lu, Y. Zhao, and Y. Xu, "An adaptive background modeling method for foreground segmentation," *IEEE Transactions on intelligent transportation systems*, vol. 18, no. 5, pp. 1109–1121, 2016.

[18] Y. Wang, Z. Luo, and P.-M. Jodoin, "Interactive deep learning method for segmenting moving objects," *Pattern Recognition Letters*, vol. 96, pp. 66–75, 2017.

[19] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognition*, vol. 76, pp. 635–649, 2018.

[20] Z. Hu, T. Turki, N. Phan, and J. T. Wang, "A 3d atrous convolutional long short-term memory network for background subtraction," *IEEE Access*, vol. 6, pp. 43 450–43 459, 2018.

[21] L. A. Lim and H. Y. Keles, "Foreground segmentation using convolutional neural networks for multiscale feature encoding," *Pattern Recognition Letters*, vol. 112, pp. 256–262, 2018.

[22] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.

[23] J. Canny, "A computational approach to edge detection," in *Readings in computer vision*. Elsevier, 1987, pp. 184–203.