Specific Vehicle Detection and Tracking in Road Environment

Yang Zhang[1 2] michael_z@live.cn
Jinqiao Wang[1] jqwang@nlpr.ia.ac.cn
Wei Fu[1] wfu@nlpr.ia.ac.cn
Hanqing Lu[1] luhq@nlpr.ia.ac.cn
Huazhong Xu[2] wutxhz@163.com
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]School of Automation, Wuhan University of Technology

ABSTRACT
In this paper, we propose a real-time method to detect and track specific vehicles, toward monitoring the abnormal activities in the traffic environment. Firstly, a novel background subtraction approach is used to get the accurate foreground segmentations and shadow suppression. Then a HIK (Histogram Intersection Kernel) based SVM classifier is trained to recognize whether a vehicle is suspicious. Finally, the Camshift based tracking is used to fast track the specific vehicles. Experiments in a real traffic scenario show the promise of the proposed approach.

Keywords
Vehicle Recognition, HIK, Background Model.

1. INTRODUCTION
Nowadays, more and more cameras are installed in public areas such as airports, highways and so on to monitor the activities in the scenes and detect abnormal. In traffic monitor environment, detecting and tracking suspicious vehicles is a critical and challenging task, which will not only help us to regulate the traffic condition better but also provide an evidence for criminal investigation by recording the whole violation process. However due to the large amount of vehicles, it is unreliable and difficult for manual operation. And traditional methods based computer vision and machine learning may also fail in some complicated scenarios. For example, different weather conditions, lighting effects including shadows and reflections, and many other factors make the problem very challenging. What's more, on-road vehicle detection and tracking have a more real-time requirement for faster processing than other application scenario since the vehicle speed is very fast especially in highway.

Recently, vision based on-road vehicle has been attracted by many researchers due to the wide application. Sun et.al presented a review of vision-based on-road vehicle detection systems in [1]. Hypothesis generation and hypothesis verification methods are introduced. For object recognition, in particular vehicle recognition, lots of approaches have been proposed in the past few years. Such as SVM based classifiers, as well as Adaboost based classifiers is popularly used in the detection task. Feris et. al [4] proposed an approach for vehicle detection in a challenge urban surveillance environment. In their work, they proposed a co-training scheme to train a detector based on appearance information.

In our application scenario, the specific vehicles include concrete mixer truck and heavy lorry in road environment. A real-time detection and tracking approach is proposed and the experimental results demonstrate the superiority.

The rest of the paper is organized as followed. In section 2 we detail our approach. We describe our method from three aspects: foreground segmentation, specific vehicle recognition and tracking. After that we give the experimental results in real road scene in section 3. Finally, conclusion of our work is given in section 4.

2. OUR APPROACH
2.1 Overview
In this paper, we propose a novel approach to detect and track specific vehicles in road environment. As shown in Fig.1, our system includes four modules: background subtraction, moving vehicle detection, specific vehicle recognition and specific vehicle tracking. We first use a method combining background subtraction and shadow suppression to extract moving vehicles. Then a classification based HIK SVM [9] is trained and to be applied in the detection phase. Finally a camera shift method is applied in our active tracking process.

2.2 Foreground Extraction
In road environment, foreground segmentation is still a challenging problem, difficulties caused by shadow and dynamic backgrounds are the key aspects. In our work, we use an efficient foreground segmentation approach to handle these difficulties. First, we use a high-efficiency background modeling approach to get a binary image which shows the moving objects [5]. Second, we subtract shadow by analyzing pixels in the Hue-Saturation-Value(HSV) color space [6] cooperate with Scale Invariant Local Ternary Pattern(SILTP) [7] to enhance the effect of shadow detecting processing. Third, we develop a Fast Connectivity Filter to remove tiny noisy blocks caused by complex scene and repair omitted blocks caused by wrong shadow subtraction.

2.2.1 Improved Background Subtraction
In most conditions, pixels in foreground moving objects changes faster than those in the background, so we can detect non-active pixels by frame to frame subtraction to confirm background pixels. Here comes the simple idea, if a pixel has been considered as a non-active pixel in continuous λ frames, then it can be finally confirmed as a real background pixel.

The frame-to-frame difference image $F_{i,j}(k)$ is used to distinguish non-active pixels and it's defined as follow:

ICIMCS'11, August 5–7, 2011, Chengdu, Sichuan, China.

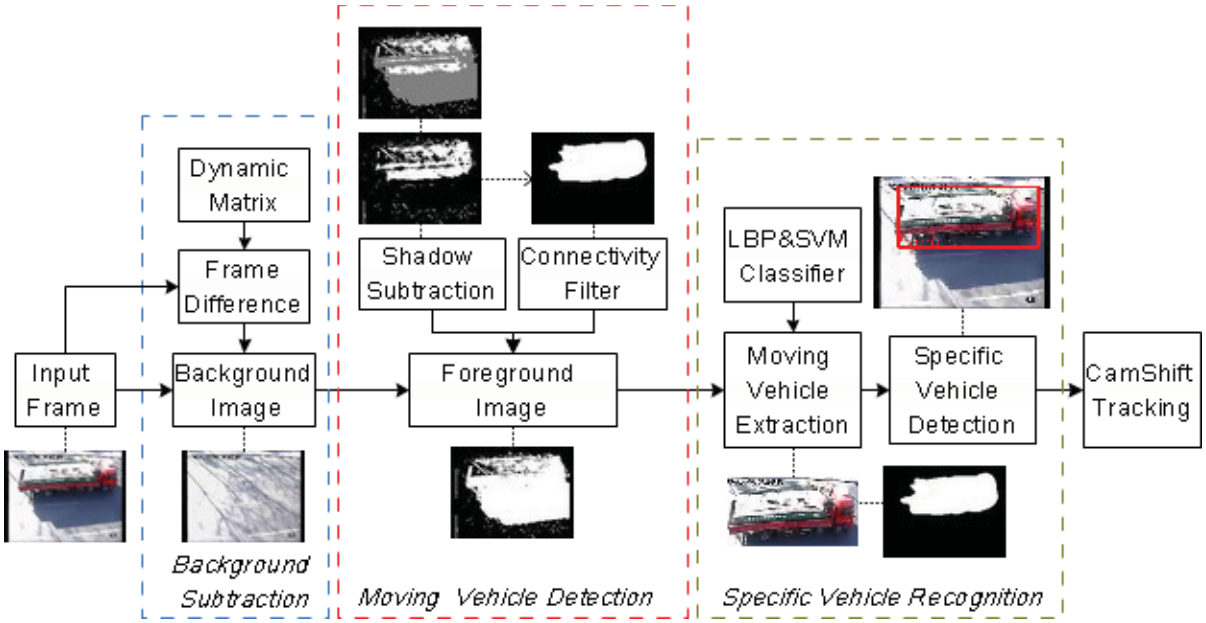


Figure.1 workflow of our approach.

$$\mathrm{F}_{i,j}(k)=\begin{cases}0 & if\ \left|\mathrm{I}_{i,j}(k)-\mathrm{I}_{i,j}(k-\gamma)\right|\le Tf\\ 1 & otherwise\end{cases} \quad (1)$$

Where $\mathrm{I}_{i,j}(k)$ represents the pixel position (i, j) in the input frame at time k, γ represents the interval time between frames which will do subtraction process. In our work $\gamma = 1$, it means we do frame subtraction once while every new frame coming. Tf is a changing threshold at time k.

In order to record and analyze the continuity of non-activeness of one pixel, we use a dynamic matrix as follows:

$$\mathrm{D}_{i,j}(k)=\begin{cases}\mathrm{D}_{i,j}(k-1)-1 & if\ \mathrm{F}_{i,j}(k)=0, \mathrm{D}_{i,j}(k-1)\neq 0\\ \lambda & if\ \mathrm{F}_{i,j}(k)\neq 0\end{cases} \quad (2)$$

Once $D_{i,j}(k)$=0, the pixel will be updated into the background with a linear model :

$$B_{i,j}(k)=\phi\cdot I_{i,j}(k)+(1-\phi)\cdot B_{i,j}(k-1) \quad (3)$$

Where $B_{i,j}(k)$ is the background image at time k and α is the weight of input frame.

Then we can get a rough segmentation image (Fig.2 c) of moving objects:

$$\mathrm{Prs}_{i,j}(k)=\begin{cases}0 & if\ MAX(\left|I_{i,j}(k)-B_{i,j}(k)\right|_c)\le Tb\\ & c=\{R,G,B\}\\ 1 & otherwise\end{cases} \quad (4)$$

Where $MAX(x_c)\ c=\{R,G,B\}$ means the maximum value of background subtraction in RGB channels and Tb is a background threshold.

2.2.2 Shadow Subtraction

In road environment, moving object segmentation with background subtraction is affected by the problem of shadows, and the area and direction of shadows are changing with The sun rose and landing. The shadows are often mistakenly classified as belonging to the foreground objects, which degrades our vehicle detection performance..

We detect shadow points (Fig.2 d) in HSV color space [6] as follow.

$$PSC_{i,j}(k)=\begin{cases}1 & if\ \alpha\le\dfrac{I_{i,j}(k).V}{B_{i,j}(k).V}\le\beta\quad \alpha,\beta\in[0,1]\\ 0 & otherwise\end{cases} \quad (5)$$

If a shadow casts on a background, the V component of shadow point in current image will be smaller than that in background image, H and S components changes within a certain limit. A preliminary sensitivity analysis for α, β is reported in [8].

This approach shows a good performance on shadow removal for kinds of vehicles, but if a car is black or it has a dark surface, large part of it will be classified as shadow points with the method above. It will lose lots of information that we can't make a complete segmentation. In order to solve this problem, we add texture analysis with Scale Invariant Local Ternary Pattern (SILTP) [7] to enhance the effect of shadow detecting processing.

$$\mathrm{PST}_{i,j}(k)=\begin{cases}1 & if\ SILTP(I_{i,j}(k).V)>Ti\\ & \wedge\left|SILTP(I_{i,j}(k).V)-SILTP(B_{i,j}(k).V)\right|>Ts\\ 0 & otherwise\end{cases} \quad (6)$$

As we know, shadows may be soft or heavy dark, soft shadows still keep the same texture feature like background, while heavy dark shadows will be pure dark gray, there won't be any texture feature.

$SILTP(I_{i,j}(k).V) > Ti$ means the local area around point (i, j) keeps some texture feature and point (i, j) might be a soft shadow point.

$\left|SILTP(I_{i,j}(k).V) - SILTP(B_{i,j}(k).V)\right| > Ts$ means the SILTP feature around point (i, j) in current image is definitely different from that in background image, so point (i, j) can't be a soft shadow point, it will be reclassified as a real moving object point. Then we get a better segmentation result (Fig.2 e):

$$PS_{i,j}(k) = \begin{cases} 1 \ \ if \ \ (\Pr s_{i,j}(k)=1 \wedge PSC_{i,j}(k) \neq 1) \\ \quad \vee (PST_{i,j}(k)=1 \wedge PSC_{i,j}(k)=1) \\ 0 \ \ otherwise \end{cases} \quad (7)$$

2.2.3 Fast Connectivity Filter

After shadow removal, the image $PS_{i,j}(k)$ has some noisy blocks in background area and small holes in moving object area, we employ a fast connectivity filter to handle the two problems.

We set a $(2N+1)^2$ size mask in image $PS_{i,j}(k)$, point (i, j) is the center position of the mask, the degree of confidence of the connectivity $C_{i,j}(k)_c$ is calculated as follow:

$$C_{i,j}(k)_c = \frac{\sum_{i-N}^{i+N} \sum_{j-N}^{j+N} PS_{i,j}(k)}{(2N+1)^2}, c \in \{f, b\} \quad (8)$$

$$\mathrm{Seg}_{i,j}(k) = \begin{cases} 1 \ \ if \ \ C_{i,j}(k)_f > C_{i,j}(k)_b \\ 0 \ \ otherwise \end{cases} \quad (9)$$

Where $\mathrm{Seg}_{i,j}(k)$ is the final moving vehicle segmentation result (Fig2.f).

The mask slips in a horizontal direction, the overlap of two adjacent positions is about $2N \times (2N+1)$ pixels, we can save the sum values of the *2N* columns respectively in a Single Linked List, when a new slipping coming, we delete the header of the linked list, calculate the sum value of the new non-overlapped column and add it to the end of the linked list, and then we can get $C_{i,j}(k)_c$ from the list. At that rate, we save a lot of computing time and make the connectivity filter run 70% faster (Fig. 3).

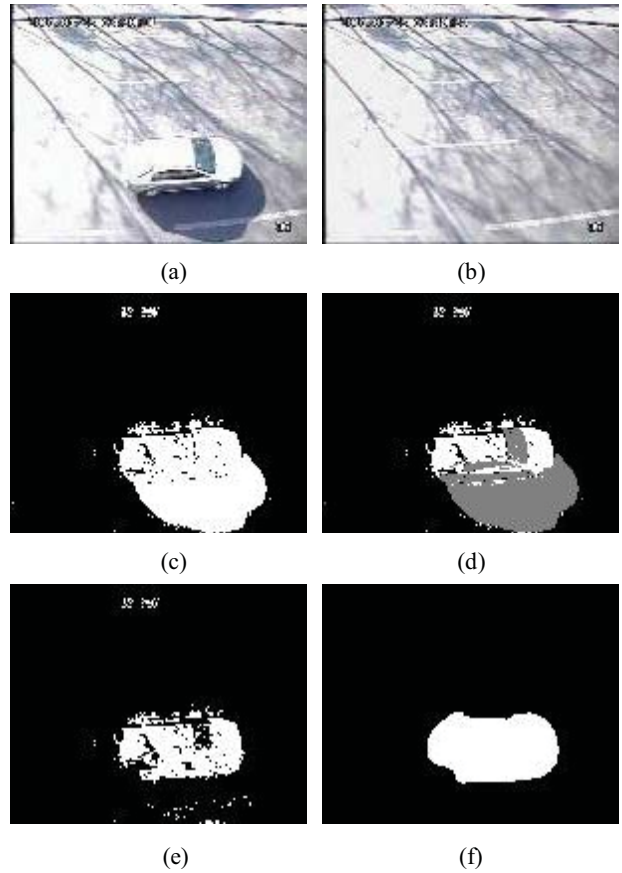


Figure.2 foreground segmentation a)original image; b) background image; c) rough Segmentation Image; d) shadow detection; e) enhance the shadow detection with SILTP; f) the result of Connectively Filter.

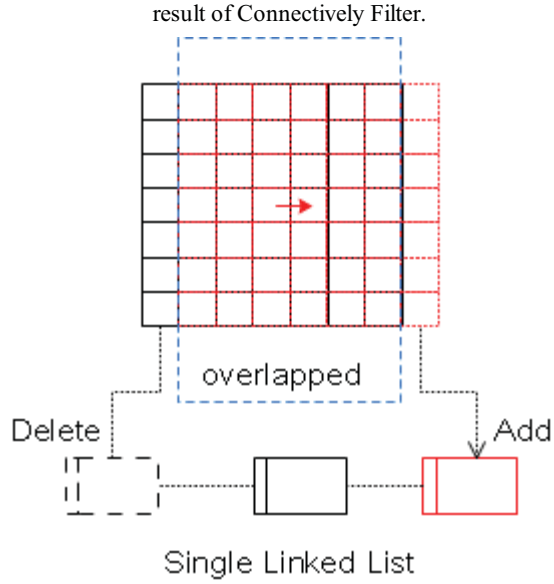


Figure.3 Schematic of Fast Connectivity Filter

2.3 Specific Vehicle Recognition

General speaking, our process of specific vehicle recognition can be divided into two stages, i.e., the offline training phase and the online detection phase.

In the training phase, the background model approach described in section 2.1 is applied to get a set of vehicle ROI images. Then after manual selection, a negative sample set and a positive sample set are manually collected for the recognizer training. The positive samples include concrete mixer truck and heavy lorry, and the negative samples are bus.

Complementally, we use area information of vehicles in $\mathrm{Seg}_{i,j}(k)$ to avoid interference of small vehicles such as cars. Then the major interference will be caused by buses, so we train buses as negative samples. Besides, it's not necessary for us to separate concrete mixer truck from heavy lorry in our application, so we train these two kinds of vehicle as positive samples to find their similar feature space.

(a)Positive Samples.

(b)Negative Samples.

Figure.4 Positive and negative samples of vehicles. The positive samples include concrete mixer truck and heavy lorry, and the negative samples are bus.

Recently histogram intersection kernel has been proved to be suitable for object recognition. In our work, we extract the LBP histogram for each training sample and then treat these features as the input of HIK SVM to learn a classifier. To be special, histogram intersection is defined as below,

$$K_{HI}(x, y) = \sum_{j=1}^{d} \min(x_j, y_j) \tag{10}$$

Where x and y are two LBP histograms with d-bins.

In the online detection phase, the foregrounds of each frame are segmented using background subtraction which is detailed in Section 2.1. And then the foreground objects are treated as the input of the SVM classifier. The specific vehicles are detected as the output of the classifier.

2.4 Specific Vehicle Tracking

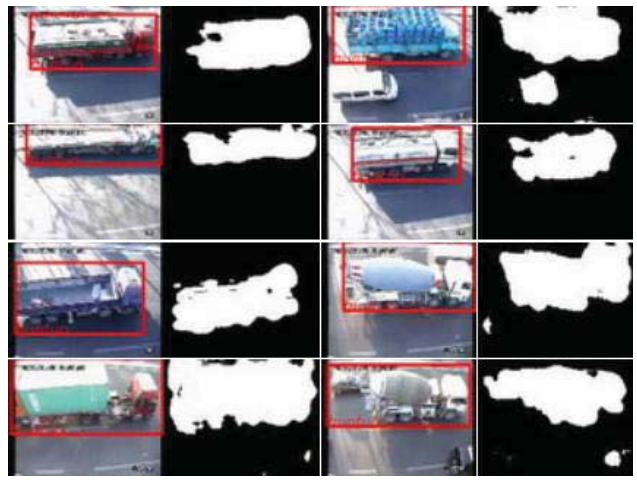
After specific vehicle detection, we utilize an active camera to track the detected vehicle. Once a specific vehicle is detected, the Continuously Adaptive Mean Shift Algorithm (CamShift) [++] is used to track it for it is fast and simple.

In practice, due to the high speed of the moving vehicle, we just track the vehicle for 2-5 seconds (after that time we think that the vehicle is out of the view) and then let the camera reset to the origin position to wait for the coming of the next specific vehicle.

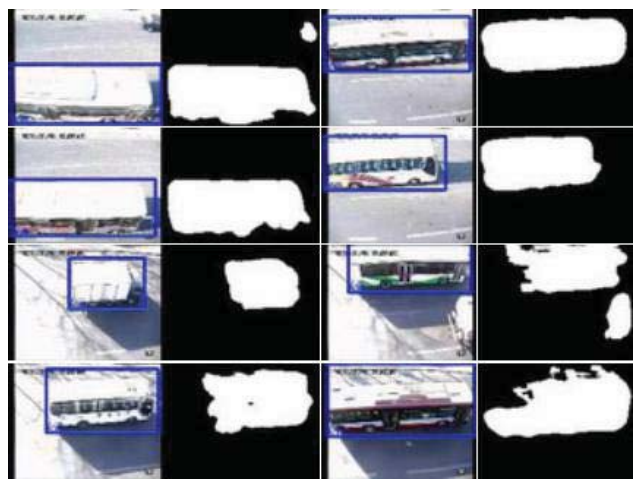
3 Experiments

In this section, we conduct our experiments on record videos from a complex road environment all day long.

Parameters for background subtraction are set as follows: $\mathrm{T}f = 15$, $\mathrm{Tb} = 25$, $\lambda = 40$, ϕ=0.08 ;Parameters for shadow suppression are set as follows: $\alpha = 0.3$, $\beta = 0.9$.



(a)Examples of specific vehicle recognition.



(b)Examples of bus recognition.

Figure.5 Experimental results of specific vehicle recognition.

The experiments are tested on a standard PC with a 2.2GHz processor and 2GB memory. The input image resolution is 160 × 120 pixels. In background modeling, the average processing time on each frame is 1.93ms. Shadow subtraction and connectivity filter costs 3.05ms and 2.25ms per frame on average, respectively. Specific vehicle recognition costs 0.19ms per frame on average. A total specific vehicle detection processing circle is 7.42ms per frame on average. Tracking costs about 6ms depend on object size. So, this system can make a good effect on real time processing.

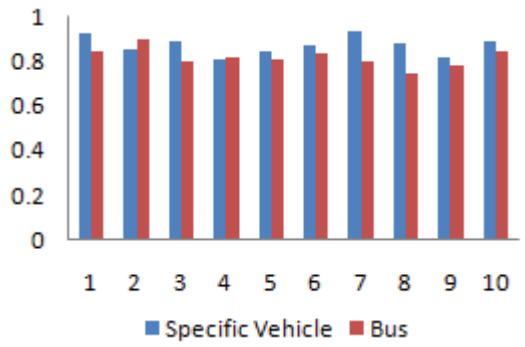


Figure. 6 Accuracy of specific vehicle recognition on test video sequences.

We test in dozens of videos which record in real road environment, the system can make a good effect on shadow suppression and the accuracy of recognition of specific vehicle is above 0.8, as shown in Fig. 6.

4 Conclusion

Detecting and Tracking specific vehicles on road is challenging topic in computer vision filed. In this paper, we propose a real-time approach for specific vehicle detection and tracking. The main contribution of our work is that an improved foreground segmentation approach is proposed, which shows excellent experimental results in the real traffic scenario. What's more, we apply a HIK based SVM classifier to recognize the specific vehicles. Then a tracking method based on CamShift is utilized to track the vehicle detected before. Finally, the whole specific vehicle monitoring system is automatic and run all the time.

5 ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China No. 60833006 and No. 60905008.

6 REFERENCES

[1] Sun, Z. H., Bebis, G. and Miller, R. 2006. On-Road Vehicle Detection: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694-711.

[2] Dalal, N. and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *CVPR*.

[3] Viola, P. and Jones, M. 2001. Robust Real-time Object Detection. In *International Journal of Computer Vision.*

[4] Feris, R., Petterson, J., Siddiquie, B., Brown, L.and Pankanti, S. 2011. Large-Scale Vehicle Detection in Challenging Urban Surveillance Environments. In *WACV*.

[5] Yang, T., Li, S. Z., Pan, Q., Li, J. 2005. Real-Time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*

[6] Cucchiara, R., Grana, C., Piccardi, M., Prati, A. 2003. Detecting Moving Objects, Ghosts, and Shadows in Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* On page(s): 1337 - 1342

[7] Liao, S. C., Zhao, G. Y., Kellokumpu V., Pietikäinen, M. and Li, S.Z. 2010. Modeling Pixel Process with Scale Invariant Local Patterns for Background Subtraction in Complex Scenes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* .

[8] Cucchiara, R., Grana, C., Piccardi, M., Prati A., and Sirotti, S. 2001. Improving Shadow Suppression in Moving Object Detection with HSV Color Information. *Proc. IEEE Int"l Conf. Intelligent Transportation Systems,* pp. 334-339.

[9] Wu, J.X. 2010. A Fast Dual Method for HIK SVM Learning. In *ECCV.*

[10] Intel Corporation (2001): Open Source Computer Vision Library Reference Manual, 123456-001.