A video-based real-time vehicle detection method by classified background learning

Xiao-jun Tan†, Jun Li†‡ & Chunlu Liu‡

Sun Yat-sen University, Guangzhou, People's Republic of China†
Deakin University, Geelong, Australia‡

ABSTRACT: A new two-level real-time vehicle detection method is proposed in order to meet the robustness and efficiency requirements of real world applications. At the high level, pixels of the background image are classified into three categories according to the characteristics of Red, Green, Blue (RGB) curves. The robustness of the classification is further enhanced by using line detection and pattern connectivity. At the lower level, an exponential forgetting algorithm with adaptive parameters for different categories is utilised to calculate the background and reduce the distortion by the small motion of video cameras. Scene tests show that the proposed method is more robust and faster than previous methods, which is very suitable for real-time vehicle detection in outdoor environments, especially concerning locations where the level of illumination changes frequently and speed detection is important.

INTRODUCTION

Video-based vehicle detection has played an important role in real-time traffic management systems over the past decade. Video-based traffic monitoring systems offer a number of advantages over traditional methods, such as loop detectors. In addition to vehicle counting, more traffic information can be obtained by video images, including vehicle classifications, lane changes, etc. Furthermore, video cameras can be easily installed and used in mobile environments.

Real-time vehicle detection in a video stream relies heavily on image processing techniques, such as motion segmentation, edge detection and digital filtering, etc. Recently, different vision-based vehicle detection approaches have been proposed by researchers [1-5]. However, two main concerns arise when they are applied to real-time vehicle detection, namely their robustness and performance. The methods should be robust enough to fit in tough outdoor environments; the algorithms should also be efficient so that the detection can be finished in time.

Recent applications of high-resolution cameras have yielded much higher performance requirements. In this article, the authors present a two level method for real-time vehicle detection that achieves the following two goals:

- *More robust*: the system can be self-adaptive to variant illumination and tolerate small motion changes caused by wind or other factors;
- *High performance*: the algorithm requires low CPU usage so that vehicles can be detected in the required time.

The article is organised as follows: a review of background subtraction is first given and followed by a proposed two-level method. Several scene tests are used in order to compare the proposed method with previous methods. There is also a discussion given in the last section.

LITERATURE REVIEW

The key technique of video-based vehicle detection belongs to a classic problem of motion segmentation. One of the widely used techniques to identify moving objects (vehicles) is background subtraction or background learning [6]. The methods fall into two categories, namely: frame-oriented methods and pixel-oriented methods. In frame-oriented methods, a predefined threshold is used to judge whether there is motion in the image scene [7]. If the diversities of the current frame and its predecessor are less than the threshold, the current frame is subtracted as the background. Those methods are easy to implement and have low CPU usage. They have been successfully applied to the detection of intruders in indoor environments, but are not practicable in traffic scenes because the illumination of outdoor environments is not stable; moreover, changes in illumination levels are regarded as motion.

In contrast, the background in a pixel-oriented method is obtained by calculating the average value of each pixel for a period much longer than the time required for moving objects to traverse the field of view [8]. By letting $I(x,y,t)$ denote the instantaneous pixel value for pixel (x,y), one can assume that the background image is the long-term average image:

$$B(x,y,t)=\frac{1}{t}\sum_{\tau=1}^{t} I(x,y,\tau) \tag{1}$$

Equation **Error! Reference source not found.** can also be computed recursively:

$$B(x,y,t)=\frac{t-1}{t}B(x,y,t-1)+\frac{1}{t}I(x,y,t) \tag{2}$$

However, this approach may fail when the lighting conditions change significantly over time. It is natural to assume that most recent images contribute more to the background than the previous image. This can be achieved by replacing $1/t$ in

Equation (2) by a constant α so that each image's contribution to the background image decreases exponentially as it recedes into the past. The background obtained by an exponential forgetting method is given by:

$$B(x,y,t) = (1-\alpha)B(x,y,t-1) + \alpha I(x,y,t)\ . \tag{3}$$

It can be shown that the exponential forgetting method is equivalent to using a Kalman filter to track the background image [9].

The average image method may produce unexpected results when objects are slow-moving or temporarily stationary. The background image becomes corrupted and object detection fails completely. Figure 1a shows the ideal background while Figure 1b is the corrupted background obtained by exponential forgetting.

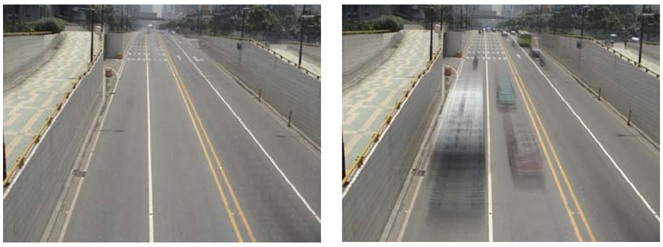

(a) (b)

Figure 1: Corrupted background image.

Several methods of adaptive background learning have recently been proposed in order to address the *slow-moving* problem by utilising variable parameters to compute the background [10][11]. These methods are generally successful in motion detection because they are based on elaborate constructed models. However, they are very time-consuming and are not suitable for real time vehicle detection scenes, where the scenes change rapidly and the processing of each frame must be finished in less than 0.1 second.

Furthermore, image changes due to small camera displacements are seldom considered and these are common in outdoor situations due to the wind load or other sources of motion that cause global motion in an image. As shown in Figure 2, some false detection can happened because of a small motion from the camera. Usually, the falsely detected pixels are distributed on the edge of the lane line and such faults can cause a detection failure in the subsequent procedure.

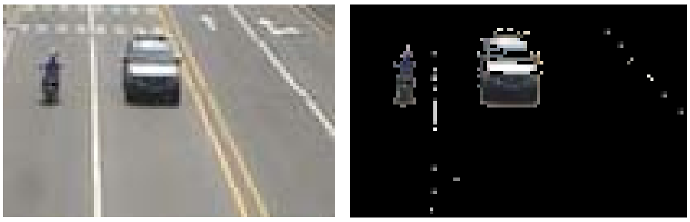

(a) (b)

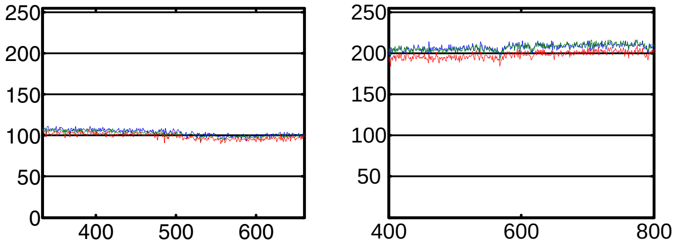
Figure 2: (a) The current frame; (b) Motion detection with some faults due to a small motion of the camera.

THE PROPOSED METHOD

A two-level method is proposed in order to address a number of issues in real world applications. The lower level of the method utilises a background learning procedure through an exponential forgetting algorithm with adaptive parameters. The higher level analyses the Red, Green, Blue (RGB) sequences of each pixel and dynamically performs pixel-classification. The parameters of the background learning procedure are determined by the higher level according to the class of the pixel.

Traffic Scene Analysis and Classification

Pixels in the traffic scene are classified into three categories according to the RGB curves, namely *road surface*, *lane line* and *others*. Figures 3a and 3b give the RGB curves of a road surface pixel and lane line, respectively. One can find two common properties from Figure 3. Firstly, the differences of three primary colours (Red, Green and Blue) are not significant. Secondly, the range of the oscillations is small. However, pixels in different categories have different average values. It is found that the road surface pixel has an average value smaller than 150 while the lane line pixel has an average value greater than 180.



(a) (b)

Figure 3: (a) Typical curves of a road surface pixel; (b) Typical curves of a lane line pixel.

Let $f(x,y,\tau)$ denote the differences of the RGB channels, and $g(x,y,\tau)$ denote the grey-scale value of the pixel located in (x,y) at the time τ, then one has:

$$\begin{aligned} f(x,y,\tau) = {} & \left[R(x,y,\tau) - G(x,y,\tau)\right]^2 \\ & + \left[G(x,y,\tau) - B(x,y,\tau)\right]^2 \\ & + \left[B(x,y,\tau) - R(x,y,\tau)\right]^2 \end{aligned} \tag{4}$$

$$g(x,y,\tau) = \frac{1}{3}\left[R(x,y,\tau) + G(x,y,\tau) + B\ (x,y,\tau)\right] \tag{5}$$

Based on the observations, a pixel classifier can be built according to the three statistical variables given by:

$$\Theta(x,y,t) = \frac{1}{N}\sum_{k=0}^{N} f(x,y,t-k) \tag{6}$$

$$\Omega(x,y,t) = S_{k=t-N}^{t}\left[g(x,y,k)\right] \tag{7}$$

$$\Psi(x,y,t) = E_{k=t-N}^{t}\left[g(x,y,k)\right] \tag{8}$$

where $S_{t1}^{t2}(\cdot)$ and $E_{t1}^{t2}(\cdot)$ are the standard deviation and average for the period $(t1,t2)$, respectively. Let T_1 and T_2 be two predefined thresholds.

Three categories are defined in this study, as follows:

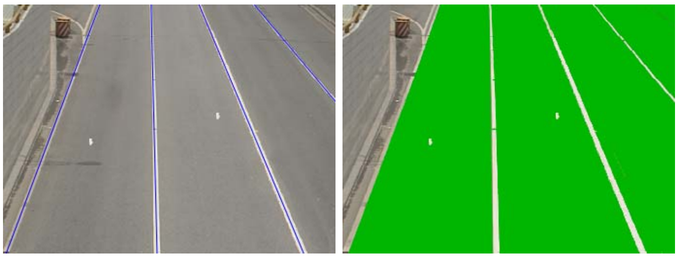
1. *Road surface*: A pixel is on the road surface if it satisfies $\Theta(x,y,t) < T_1$, $\Omega(x,y,t) < T_2$ and $\Psi(x,y,t) < 150$;

2. *Lane line*: A pixel is on the road surface if it satisfies $\Theta(x,y,t) < T_1$, $\Omega(x,y,t) < T_2$ and $\Psi(x,y,t) > 180$;
3. *Others*: The pixels in this category are neither on the road surface nor on the lane line.

The two thresholds, T_1 and T_2, can be understood as detailed below. Firstly, T_1 is the threshold of colour saturation. Both the road surface and land separators are very low in colour saturation. Secondly, T_2 is the threshold of stability, which is used to solve the problem of small camera motions caused by the wind load or other reasons. As illustrated in Figure 2, there are some faults at the edge of the road surface and land separators due to the small motion of the camera. By setting the appreciate threshold T_2, these edge pixels are regarded as *others*, rather than *road surface* or *lane lines*. In the experiments elaborated on in this article, T_1 and T_2 have been experimentally chosen. In the case, T_1 is set to 10.0 and T_2 is set to 5.0.

After the classification, a procedure is added to enhance the robustness of the classification, which is based on the observations of the geometric characteristics of the highway. Firstly, a lane line is some kind of straight line. Secondly, the road surface is an area of connective patterns. The robustness enhancements can be carried out as follows:

- *Line Detection*: With the results of the classifier, the candidates of a lane line form a straight line, which can be detected using a Hough Transformation [2]. Only those near the line can be confirmed as lane lines. Figure 4a shows the detected line in the colour blue. Two white areas due to plastic bags are excluded. If the lane line is another curve than a straight line, then other techniques can be adopted [12][13];
- *Pattern Connectivity*: A growing algorithm, similar to that from Tremeau and Borel, is utilised to check the candidates of the road surface [14]. The result is shown in green in Figure 4b. In Figure 4a, the blue lines indicate the detected lines of lane lines. And in Figure 4b, the green pixels are the connected patterns of road surfaces. Those pixels that cannot be classified – either as road surface or as a lane line – are labelled with their original colour.



(a) (b)

Figure 4: The result of a more robust classification.

Background Learning based on Classification

The background-learning module of the proposed method has the similar form as the *exponential forgetting* algorithm mentioned above:

$$B(x,y,t) = (1-\hat{\alpha})B(x,y,t-1) + \hat{\alpha}I(x,y,t) \qquad (9)$$

However, the forgetting factor, $\hat{a}$, is here no longer a fixed constant, but rather a function of the pixel value (x,y,t), for which $\hat{a}$ is given by:

$$\hat{\alpha} = \frac{1}{[I(x,y,t) - I(x,y,t-\Delta t)]^2 + \beta}. \qquad (10)$$

where β is the category parameter. If a pixel belongs to the road surface, then β = 20 and $\hat{a}$ has a maximum value of 0.05. If the pixel belongs to the lane line, then β = 10 and $\hat{\alpha}$ has a maximum value of 0. The factor $\hat{a}$ is set according to the statistical models of the different class of traffic scene. The advantage is that the noise can be suppressed differently.

When compared with the simple *forgetting* algorithm, the modified forgetting factor $\hat{a}$ considers the motion happening at the current pixel so that only the relatively *stable* pixels are chosen to perform background learning. This approach can significantly reduce the effects caused by small motion of video cameras, which is very important for outdoor environments, especially mobile environments.

If the pixel belongs to the *other* class, then background learning will not be performed because this kind of pixel has nothing to do with the motion detection of the current frame, as detailed in the next section. This can reduce the calculation.

To further reduce the computation time, the background classification and background learning in the proposed method are not updated synchronously. As shown in Figure 5, background learning is performed for almost every frame, but the classification will only be carried out after a certain period. Although it takes a relatively long time to undertake the classification, the real-time detection is not affected.

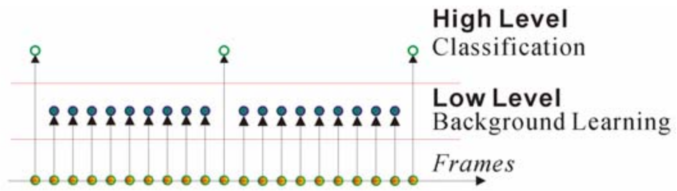


Figure 5: Frequency of the classification and background learning.

Vehicle Detection

Vehicle detection has almost the same meaning of motion segmentation, which judges whether there is a motion at a certain pixel. In the proposed method, the area to detect vehicles is not the whole image, but rather due to understanding given at the higher level. If the pixel belongs to the *other* class and is in the lane line or *near* the lane line, then detection will not be performed. This can help to avoid false detection due to small motions of the camera.

Figure 6 shows each area with a possible vehicle, while other areas are ignored. An algorithm named Minimum Boundary Rectangles (MBRs) is then utilised in order to divide those areas that belong to different vehicles [1]. In Figure 6, the rectangles indicating the moving vehicles are outlined.

EXPERIMENTAL RESULTS

Three different traffic scenes are selected to test the proposed method and a comparison is made with two other background learning approaches, namely *Average Image* and *Adaptive Method*. The scenes are taken from three highways and all videos are taken during daytime with a fixed camera with the resolution set at 360×288. The three scenes are as follows:

- Scene 1: There is no slow-moving vehicle and no small motion of the fixed camera;
- Scene 2: There is a small motion caused by the wind; the maximum displacement is set to four pixels;
- Scene 3: Slow-moving vehicles can be observed frequently.

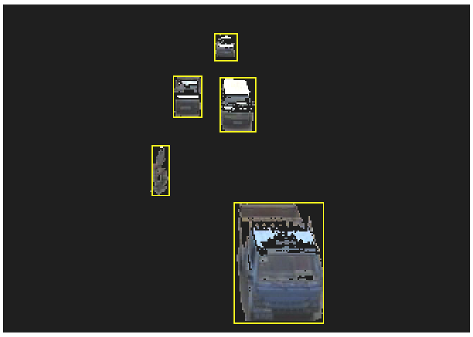


Figure 6: Vehicle detection using the MBR algorithm.

The programs are written in C++ and executed on a Pentium IV (1.8GMHz) PC, with the format of each video frame being 24bit RGB pixels. As the main concerns are the robustness and performance, two comparisons are given, namely the recognition rate and the computing time. The results are shown in Tables 1 and 2.

Table 1: Recognition rate using different methods (%).

	Scene 1	Scene 2	Scene 3
Average Image	81	51	59
Adaptive Method	98	71	74
Proposed method	98	94	95

Table 2: Average processing time and frame rates.

	Processed Frames/s	Average Processing Time (ms)
Average Image	7.39	136
Adaptive Method	1.17	855
Proposed method	10.01	100

One can find that the proposed method is the fastest, while adaptive method is the slowest. The proposed method detects vehicles with a success rate of more than 90% in all three scenes; the other two methods have significantly lower success rates for Scenes 2 and 3. These tests have shown that the proposed method is more robust and faster than the other two methods.

CONCLUSIONS

In this article, the authors present a two-level method to fit in the needs of real-time video-based vehicle detection. There are two improvements over previous methods. The background pixels are classified according to the RGB curves. The criterion based on background pixel characteristics can significantly reduce the fault judgements caused by a small motion of the video cameras. The robustness of the method is further enhanced by the geometric characteristics of the lane line and road surface. These classifications also make it possible to define different adaptive parameters for different background pixel type. To achieve increased efficiency, the background classification is not updated synchronously with background learning. Such an approach significantly reduces the computer power required by the algorithm, while the robustness of the algorithm is retained. The experimental tests show that the proposed method is more robust and efficient than previous methods.

The proposed method is very suitable for real-time vehicle detection applications that require efficient algorithms to process video images in time and with strong robustness in order to obtain reliable results. However, it should be noted that the proposed method is not suitable for vehicle detection at night when there is insufficient illumination. Finding a suitable classification method and optimal parameters for night scenes is open for future study.

REFERENCES

1. Kim, J.B. et al, Wavelet-based vehicle tracking for automatic traffic surveillance. *Proc. IEEE Region 10 Inter. Conf. on Electrical and Electronic Technology*, Singapore, 1, 313-316 (2001).
2. Lee, J.W. et al, A study on recognition of lane and movement of vehicles for port AGV vision system. *Proc. 2002 IEEE Inter. Symp. on Industrial Electronics*, L'Aquila, Italy, 2, 463-466 (2002).
3. Chen, S.C. et al, Spatiotemporal vehicle tracking – the use of unsupervised learning-based segmentation and object tracking spatiotemporal vehicle tracking. *Robotics & Automation Mag*., 12, **1**, 50-58 (2005).
4. Wang, Y.K. and S.H. Chen. Robust vehicle detection approach. *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, Como, Italy, 117-122 (2005).
5. Xie, L. et al, Real-time vehicles tracking based on Kalman filter in a video-based ITS. *Proc. 2005 Inter. Conf. on Communications, Circuits & Systems*, Hong Kong, China, 2, 883-886 (2005).
6. Friedman, N. and Russell, S., Image segmentation in video sequences: a probabilistic approach. *Proc. 13th Conf. on Uncertainty in Artificial Intelligence*, Providence, USA (1997).
7. Kuno, Y. et al, Automated detection of human for visual surveillance system. *Proc. Inter. Conf. on Pattern Recognition*, 3, Vienna, Austria, 865-869 (1996).
8. Ivanov, Y., Bobick, A. and Liu, J., Fast lighting independent background subtraction. *Inter. J. of Computer Vision*, 37, **2**, 199-207 (2000).
9. Koller, D., et al. Towards robust automatic traffic scene analysis in real-time. *Proc. 12th IAPR Inter. Conf. on Pattern Recognition*, Jerusalem, Israel, 1, 126-131 (1994).
10. Huwer, S. and Niemann, H., Adaptive change detection for real-time surveillance applications. *Proc. 3rd IEEE Inter. Workshop on Visual Surveillance*, Dublin, Ireland, 37-46 (2000).
11. Stauffer, C. and Grimson, W.E.L., Adaptive background mixture models for real-time tracking. *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Fort Collins, USA, 2, 246-252 (1999).
12. Enkelmann, W., Video-based driver assistance – from basic functions to applications. *Inter. J. of Computer Vision*, 45, **3**, 201-221 (2001).
13. Guiducci, A., Camera calibration for road applications. *Computer Vision and Image Understanding*, 79, **2**, 250-266 (2000).
14. Tremeau, A. and Borel, N., A region growing and merging algorithm to color segmentation. *Pattern Recognition*, 30, **7**, 1191-1203 (1997).