Video-based traffic data collection system for multiple vehicle types

Shuguang Li1, Hongkai Yu1, Jingru Zhang1, Kaixin Yang2, Ran Bin2

1School of Electronic and Control Engineering, Chang’an University, Xi’an 710064, People’s Republic of China
2Department of Civil and Environmental Engineering, University of Wisconsin, Madison 53705, USA

E-mail: lxslsg@vip.sina.com

Abstract: Traffic data of multiple vehicle types are important for pavement design, traffic operations and traffic control. A new video-based traffic data collection system for multiple vehicle types is developed. By tracking and classifying every passing vehicle under mixed traffic conditions, the type and speed of every passing vehicle are recognised. Finally, the flows and mean speeds of multiple vehicle types are output. A colour image-based adaptive background subtraction is proposed to obtain more accurate vehicle objects, and a series of processes like shadow removal and setting road detection region are used to improve the system robustness. In order to improve the accuracy of vehicle counting, the cross-lane vehicles are detected and repeated counting for one vehicle is avoided. In order to reduce the classification errors, the space ratio of the blob and data fusion are used to reduce the classification errors caused by vehicle occlusions. This system was tested under four different weather conditions. The accuracy of vehicle counting was 97.4% and the error of vehicle classification was 8.3%. The correlation coefficient of speeds detected by this system and radar gun was 0.898 and the mean absolute error of speed detection by this system was only 2.3 km/h.

1 Introduction

The proportion of large vehicles (LVs) affects the interactions among vehicles and the efficiency of roadway. Highway capacity manual [1] requires LV data for the analysis of traffic conditions, traffic safety and pavement design. Therefore traffic data of multiple vehicle types are important for the research of traffic flow.

Owing to the complexity of the mixed traffic conditions, induction loops are not useful to collect such data. Most of existing video-based detectors can only provide several macroscopic traffic parameters such as flow and mean speed without vehicle classification. Many of these detectors are sensitive to road reflections, illumination changes, shadow interferences etc. These detectors always ignore cross-lane vehicles and may repeatedly count one vehicle, which influences their accuracy of vehicle counting. Their accuracy of classification is obviously affected by vehicle occlusions. Many of these detectors are insensitive to road reflections, illumination changes, shadow interferences etc. These detectors always ignore cross-lane vehicles and may repeatedly count one vehicle, which influences their accuracy of vehicle counting. Their accuracy of classification is obviously affected by vehicle occlusions.

A new video-based traffic data collection system for multiple vehicle types is developed in this paper. This system tracks every passing vehicle for several frames, and obtains the type and speed of every passing vehicle. This system combines the vehicle type with the vehicle speed and outputs classified vehicle data including flows and mean speeds of multiple vehicle types. A colour image-based adaptive background subtraction is proposed to improve the reliability of object detection. A series of processes like shadow removal and setting road detection region are used to improve the system robustness. The cross-lane vehicles are taken into consideration and the repeated counting for one vehicle is avoided, which significantly improves the accuracy of vehicle counting. The space occupation of the blob and data fusion are introduced to reduce the classification errors caused by vehicle occlusions.

2 Previous work

The video-based traffic data collection has been an area of interest in intelligent transportation system for the past few decades. Michalopoulos et al. [2] introduced the useful autoscope system which has been widely used for traffic data collection. Early video detection systems are effective for collecting macroscopic traffic parameters such as flow, mean speed and density. Further researches about vehicle classification and collecting microscopic traffic characteristics emerge.

Former studies by Li et al. [3, 4] introduced a video-based collection system for multi-type vehicles’ traffic data. Yuan et al. [5] performed a computer vision system composed of five models for vehicle classification including length-based classification. The experiments showed that some geometric parameters of vehicles such as length, width and height are useful for classification. Gupte et al. [6] used regions to detect, track and classify vehicles. However, the accuracy was affected by occlusions and shadows. Zhang et al. [7]
placed a virtual detection loop on each lane to extract the pixel-based vehicle length and simply divided the vehicles into long vehicles and short vehicles. Unfortunately, their system cannot detect the cross-lane vehicles and it was insufficient for classification by only using vehicle length because of the errors caused by longitudinal occlusions. Vehicle occlusions are frequent in the video-based traffic data collection system, which exert a detrimental impact on the system accuracy [8]. In order to solve the problems caused by occlusions and improve the detection accuracy, some researchers take advantage of more detailed features of vehicles and advanced pattern recognition methods to classify and track vehicles. Corner features were computed to reduce the errors caused by occlusions [9, 10], however their studies had the following limitations: (i) Corner features do not cover the whole vehicle, so the position and geometric parameters obtained from the detected vehicle may have some mistakes. (ii) The missed corner features may result in apparent errors for speed detection. Ozkurt et al. [11] presented a feed-forward neural network to classify vehicles, but neural network needs time-consuming training and requires a lot of prior data. Hasegawa and Kanade [12] demonstrated a vision system that can classify vehicles and estimate the main colour of targets. However it needed to collect image samples for learning to classify vehicles. Zhang et al. [13] developed a vehicle classification framework based on principal component analysis, however the accuracy of such algorithm depended on the training vehicle samples. Mallikarjuna et al. [14] computed a five-dimensional feature vector for classification and used a temporal tracking methodology for tracking. The approach not only obtained several macroscopic parameters but also some microscopic parameters, however its computation was complex and hard for real-time traffic data collection. Khan et al. [15] presented an approach that uses detailed three-dimensional (3D) models to detect and classify vehicles, however it needed to build 3D matching vehicle models first and its accuracy was obviously affected by the training 3D models. Kafai et al. [16] used a dynamic Bayesian network to classify vehicles, however this method based on the rear-side view where the camera’s field of view was directly behind the vehicle and therefore was limited for the common usage.

In general, these previous researches still have the following shortcomings: (i) The robustness and reliability of object detection is not stable. (ii) Some detection systems ignore detecting cross-lane vehicles and may repeatedly count one vehicle. (iii) The accuracy of classification is obviously affected by vehicle occlusions. (iv) Some classification methods need to collect image samples for learning to classify vehicles, so the classification is obviously affected by the environment and shooting style.

The system described in this study makes efforts to overcome these shortcomings. First, a colour image-based adaptive background subtraction is proposed to improve the reliability of object detection and a series of processes like shadow removal and setting road detection region are used to improve the system robustness. Second, some cross-lane virtual detection loops are set up to detect the cross-lane vehicles and a special counting limit is used to avoid the repeated counting. Third, the space occupation of the blob and data fusion are used to reduce the classification errors caused by vehicle occlusions. Finally, the geometric parameters of vehicles are exploited for classification, so the classification method is universal and does not need training and learning.

3 Methodology

In this section, the major algorithms of the system are described.

3.1 System overview

Camera is located directly above the target road with its optical axis tilting a certain angle downward. The one-way traffic on multiple lanes is monitored at the same time. The system consists of six modules: user input, initialisation, colour image-based adaptive background subtraction, vehicle counting, vehicle tracking and data fusion. The vehicle tracking module has two sub modules: speed detection and vehicle classification. Fig. 1 shows the flow chart of the system and the green parts in this figure show the differences and innovations of this system.

3.2 User input

Before starting the system for traffic data collection, first users must input some basic parameters: the detailed positions and sizes of virtual detection loops, the classification thresholds and the shadow removal thresholds.

The detection accuracy is related to the specific positions and sizes of virtual detection loops. The virtual detection loops should be placed in locations where vehicles are clearly visible with rare occlusion problems. The width of every virtual detection loop is a little smaller than the width of one lane. Fig. 2 illustrates the positions and sizes of the virtual detection loops. In order to avoid missing vehicles, user should place a virtual detection loop on every lane (Loop1, Loop2 and Loop3 in Fig. 2) and also place some cross-lane virtual loops (Loop4 and Loop5 in Fig. 2) to detect cross-lane vehicles.

3.3 Initialisation

The initialisation module consists of initial colour background extraction, setting road detection region and traffic scene calibration.

It is important to extract the accurate initial background for background subtraction and vehicle detection. The initial colour background is obtained by calculating the average value of each pixel from a collection of colour images. The average value of each pixel is calculated in the every channel of RGB colour space for the starting N frames, and the initial colour background is obtained by constructing a colour image using these average values in three channels. Figs. 3a and b show a snapshot of a video scene and the extracted initial colour background.

After obtaining initial background, the system uses the initial background to set the road detection region. The Canny edge detection method and Hough transform are then utilised to detect road edge lines. Fig. 3c shows the detected road edge lines, which are drawn in red. After detecting road edge lines, the system automatically sets up the road detection region. Fig. 3d shows the road detection region defined by system and the uninterested region is removed. All the following operations and calculations are processed in the road detection region to reduce the calculation complexity and interferences.

For accurate speed detection, the positions of vehicles in the image should be transformed to their real-world coordinates, which is a 2D-to-3D mapping. Direct linear transformation does not need to calculate the intrinsic and
extrinsic matrixes of the camera and directly builds the linear transformation model to transform the image coordinates to real-world coordinates. So direct linear transformation is used for traffic scene calibration. More details about direct linear transformation are described in [17].

3.4 Colour image-based adaptive background subtraction

Nowadays, grey image-based background subtraction is widely used in many video-based detection systems. However, grey image-based background subtraction has the following shortcomings. If the grey scale of the vehicle is similar with that of the background, grey image-based background subtraction fails to fully detect the vehicle. Some vehicles which have different colours may have the similar brightness with the road, so the grey scale of these vehicles is similar with the grey scale of the road in the image. Consequently, these vehicles cannot be segmented completely and detected well by grey image-based background subtraction. In order to reduce this error, colour image-based adaptive background subtraction is used to detect vehicles.

The system has obtained the initial colour background in the module of initialisation. In order to adapt to light
changes and the complex traffic environment, the colour background should be updated periodically. The update algorithm for dynamic background presented by Gupte et al. [6] is used to update background in the every channel of RGB colour space. Then the dynamic colour background is synthesised by every updating background in three channels.

A difference image can be obtained by subtracting the dynamic colour background in the red channel, the green channel and the blue channel, respectively. The resultant difference image is then segmented by an adaptive threshold to get binary objects. Threshold segmenting is the key part of object detections. The static threshold cannot be used to rightly compute binary objects because of changing traffic conditions. Therefore an optimal Otsu algorithm [18] is calculated to get the adaptive threshold to detect binary objects after colour image-based background subtraction. Some morphology operations such as dilatation operations and close operations are then applied to the detected binary objects and the inside holes of the binary objects are filled, which significantly strengthens the vehicle objects.

**Fig. 2** Main user interface of the system

**Fig. 3** Initialisation

a A snapshot of a video scene  
b Extracted initial colour background  
c Road edge lines detection by Hough transform  
d Road detection region defined by system
Fig. 4 shows the differences of object detections between colour image-based adaptive background subtraction and grey image-based adaptive background subtraction. Fig. 4 shows that colour image-based adaptive background subtraction obtains more accurate and complete vehicle objects and improves the system reliability. It is because that colour image-based adaptive background subtraction makes a synthesis of every channel of RGB colour space.

3.5 Vehicle counting

3.5.1 Shadow removal: The vehicle shadow moves with the vehicle, so the shadow may be incorrectly judged to be the foreground object. Shadows cause some serious problems such as wrong foreground objects, horizontal occlusions of vehicles and misclassifications. To reduce the errors caused by shadows, a HSV colour space-based shadow detection algorithm proposed by Cucchiara et al. [19] is used to detect and remove the shadows. The current colour image in RGB colour space is transformed to HSV colour space, and it is compared with the dynamic colour background for each pixel by using a judgment rule described by this algorithm. If a pixel is judged to be shadow, this pixel is removed in the foreground image.

3.5.2 Vehicle counting: In the foreground image, the change percentage of the average grey value of the virtual detection loop in the consecutive frames is monitored to detect vehicles. Each virtual detection loop has one independent counter. The change percentage can be obtained as follows

\[
\Delta MGP = M_{G_i} - M_{G_{i-1}} \tag{1}
\]

\[
M_G = \frac{1}{n} \sum_{(x,y) \in Loop} I(x, y) \tag{2}
\]

where \(\Delta MGP\) is the change percentage of the average grey value of the virtual detection loop in the consecutive frames; \(M_G\) is the average grey value of the virtual detection loop; \(M_{G_i}\) and \(M_{G_{i-1}}\) represent the average grey value of the virtual detection loop in the frame \(i\) and the frame \(i - 1\), respectively; \(n\) is the area of the virtual detection loop in number of pixels; \(I(x, y)\) is the foreground image:

\[
I(x, y) = \begin{cases} 
1, & \text{foreground} \\
0, & \text{otherwise} 
\end{cases} \tag{3}
\]

If the change percentage \(\Delta MGP\) is greater than some threshold \(\Delta T_v\) (in the system, 10% is used), it demonstrates that one vehicle is passing. Then the corresponding counter is started at the same time.

3.5.3 Avoid repeatedly counting: Some binary vehicle objects contain big holes inside in the foreground image because of poor segmentations, so \(\Delta MGP\) increases to be greater than \(\Delta T_v\) for several times causing repeated count. Some vehicles are so large that it may also result in the repeated count by the adjacent virtual detection loops. The accuracy of video-based traffic systems is always damaged by repeatedly counting. So a special limit is set to avoid repeatedly counting in this paper.

One vehicle needs some time to pass through the detection loop and the following vehicle also needs some time to arrive at the detection loop. There are some spaces among different vehicles, so some adjacent detection loops cannot be occupied at the same time. For example, if Loop2 in Fig. 2 is occupied by one vehicle, Loop4 and Loop5 will not be occupied by vehicles at the same time. The adjacent detection loops which cannot be occupied by vehicles at the same time are called adjacent conflict loops (ACL) in this paper. A special limit for avoiding repeatedly counting is shown in Fig. 5.

If one detection loop LoopX has detected one vehicle, the corresponding counter adds and the system records the current frame \(F_c\). From the current frame \(F_c\) on, LoopX and its ACL all stop detection and counting in the next \(F_s\).
3.6 Vehicle tracking

One vehicle is constantly tracked for M frames after it is detected by one virtual detection loop (M ≤ 15). During the tracking procedure, the system detects the speed of the tracked vehicle and classifies the tracked vehicle at the same time.

3.6.1 Vehicle tracking algorithm: Various computer vision-based algorithms can be used to track vehicles: feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.

The vehicle motion is small in one frame interval, so the feature-based tracking, mean shift-based tracking, Kalman filter-based tracking and so on [20]. Feature-based tracking which is simple and fast can meet the real-time requirements, so the feature point of vehicle head (FPVH) for tracking.
area and length of the vehicle. The errors caused by occlusions are more apparent when the camera is not located very high. When the camera is located directly above the target road section, the vehicle occlusions have three conditions: horizontal occlusion, slant occlusion and longitudinal occlusion, which are shown in Fig. 6. The horizontal occlusion shown in Fig. 6a is mainly caused by shadows, so the horizontal occlusion is few remaining after shadow removal. The slant occlusion and longitudinal occlusion are more common.

To reduce the misclassifications caused by the slant occlusion shown in Fig. 6b, the space occupation of the blob is used to detect the slant occlusion. In this study, blobs represent the vehicles in the foreground image. The space occupation of the blob is defined as follows

\[ R = \frac{S_b}{S_r} \]  

(9)

where \( S_b \) is the pixel-based area of the blob in the image coordinate axis; and \( S_r \) is the pixel-based area of the blob’s minimum enclosing rectangle in the image coordinate axis. The minimum enclosing rectangle of the blob has been drawn in Fig. 6.

The area and length of the blob is bigger if slant occlusion happens, so the blob may be incorrectly judged as a LV. The space occupation of the blob, \( R \), is lower if the slant occlusion happens. However, the space occupation of the blob, \( R \), is higher if the blob is a LV. So it is feasible to set up a threshold of the space occupation of the blob to reduce the errors caused by the slant occlusion.

The errors caused by the longitudinal occlusion shown in Fig. 6c are hard to reduce for the video detection system. The inter-vehicle gap is required to be enough big to distinguish two closely spaced vehicles on a level roadway [21]. Owing to the limit of the camera shooting angle, the longitudinal occlusion happens if the relative distance between two vehicles in the longitudinal direction is small. Therefore it is not accurate to classify the vehicle in a single frame. The system constantly tracks the vehicle for \( M \) frames and the vehicle type is judged every tracking frame. Finally, the \( M \) classification results are synthesised to determine the final type of the tracked vehicle.

In this system, the vehicles are divided into two types: LV and small vehicle (SV). The pixel-based length, the pixel-based area and the space occupation of the tracked vehicle are extracted. The system uses these geometric parameters to classify the vehicle. The method of classification in every tracking frame is expressed as follows

\[
\begin{align*}
\text{type}_t &= \begin{cases} 
LV & \text{if } L > l_1 \text{ and } S > s_2 \text{ and } R > r_1 \\
SV & \text{otherwise}
\end{cases} \\
\forall t &\in [1, M]
\end{align*}
\]  

(10)

where \( L, S, R \) are the pixel-based length, the pixel-based area and the space occupation of the tracked vehicle, respectively; and \( l_1, s_1 \) and \( r_1 \) are the classification thresholds, respectively.

### 3.7 Data fusion

One vehicle is constantly tracked for \( M \) frames, so the \( M \) instantaneous speeds and vehicle types are obtained for the tracked vehicle. Then data fusion is used to determine the final speed and vehicle type of the tracked vehicle.

#### 3.7.1 Final speed for the tracked vehicle

The final speed of the tracked vehicle, \( V \) (km/h), can be...
expressed as follows

\[ V = 3.6 \times \left( \frac{1}{M} \sum_{i=1}^{M} V_i \right) \]  

(11)

where 3.6 is a factor to covert m/s into km/h.

**3.7.2 Final vehicle type for the tracked vehicle:** A vote algorithm is used to determine the final type of the tracked vehicle. According to the majority principle, the system votes for the final vehicle type. The final vehicle type of the tracked vehicle, type, is determined as follows

\[
\text{type} = \begin{cases} 
\text{LV} & \text{if } N_{LV} > N_{SV} \\
\text{SV} & \text{otherwise}
\end{cases} \quad N_{LV} + N_{SV} = M \quad (12)
\]

where \(N_{LV}\) and \(N_{SV}\) are the detected numbers of LV and SV in the \(M\) instantaneous vehicle types, respectively.

Data fusion reduces the stochastic noises of speed detection and overcomes the difficulties and errors to classify vehicles in one single frame. It is worth mentioning that data fusion is able to reduce the misclassifications caused by vehicle occlusions especially the longitudinal occlusions.

**3.8 Vehicle tracking display**

Fig. 7 demonstrates the vehicle tracking for constant \(M\) frames. Fig. 7 shows only four frames during the tracking procedure. In the Fig. 7a, one virtual detection loop detects a vehicle and the minimum enclosing rectangle and the length of this vehicle are drawn in green at the same time. This system starts to track the vehicle. The vehicle is then tracked for constant \(M\) frames, and vehicle classification and speed detection are processing every tracking frame. The minimum enclosing rectangle and vehicle length of this vehicle are drawn in red during the tracking procedure, which is shown in Figs. 7b and c, respectively. Finally, the final speed and type of the tracked vehicle are given, which is shown in Fig. 7d. Fig. 7 also shows that the detection and tracking of different virtual detection loops are independent and multiple vehicles can be tracked, respectively, at the same time.

**4 Experiment results**

To verify the accuracy of the system, four offline traffic videos under different weather conditions were used to test it. The four test videos were all recorded in the Middle part of South Second Ring (city expressway) of Xi’an, China and the time length of each video was ten minutes. The test video A was recorded at 3:00 pm on 15th March 2011, and the test condition was sunny with intense shadows. The test video B was recorded at 4:00 pm on 20th March 2011, and the test condition was sunny with weak shadows. The test video C was recorded at 1:00 pm on 10th July 2011, and the test condition was cloudy without shadows. The test video D was recorded at 3:00 pm on 9th June 2011, and the test condition was drizzle with relatively wet road surface. The radar gun was used to detect the vehicle speeds on one lane during the camera shooting procedure, then the speeds detected by this system and radar gun were compared.

Table 1 shows the comparisons between the results of vehicle counting and classification by the system and ground-truth data. Table 1 concludes that the accuracy of vehicle counting was 97.4%. The errors of vehicle
classification without data fusion was 18.8% and the errors of vehicle classification using data fusion was 8.3%, which can be seen in Table 1 as well. Vehicle occlusions result in over-counted LVs and lead to serious vehicle classification errors. Table 1 shows that data fusion can reduce these errors. The experiment results show the accuracy and effectiveness of the vehicle counting and classification of the system.

The speed detected by radar gun is used as the ground-truth of the vehicle speed. Fig. 8 shows the comparisons between speeds detected by this system and radar gun. Fig. 9 shows the histograms of the errors of speeds detected by this system.

The average correlation coefficient of speeds detected by these two ways was 0.898 and the mean absolute error of speeds detected by this system was 2.3 km/h. Fig. 9 shows that the error distribution of the speed estimate of this system is nearly normal distribution. Therefore the speeds detected by this system are reliable.

Finally, the system outputs the traffic data of multiple vehicle types. Take the test video A for example, the system output the traffic data as follows: total flow: 4632 vehicle/h; flow of LVs: 90 vehicle/h; flow of SVs: 4542 vehicle/h; mean speed of total vehicles: 44.2 km/h; mean speed of LVs: 41.5 km/h; mean speed of SVs: 44.3 km/h.

### Table 1 Vehicle Counting and Classification

<table>
<thead>
<tr>
<th>Test Video</th>
<th>Vehicles, G</th>
<th>Vehicles, S</th>
<th>Vehicle counting accuracy, %</th>
<th>LV, G</th>
<th>Errors of LV counting (Scwdf)</th>
<th>Errors of classification (Scwdf), %</th>
<th>Errors of LV counting (Scudf)</th>
<th>Errors of classification (Scudf), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>793</td>
<td>772</td>
<td>97.4</td>
<td>16</td>
<td>1 a, 2 b</td>
<td>18.8</td>
<td>1 a</td>
<td>6.3</td>
</tr>
<tr>
<td>B</td>
<td>760</td>
<td>751</td>
<td>98.8</td>
<td>14</td>
<td>1 a, 1 b</td>
<td>14.3</td>
<td>1 a</td>
<td>7.1</td>
</tr>
<tr>
<td>C</td>
<td>622</td>
<td>615</td>
<td>98.9</td>
<td>10</td>
<td>1 a, 1 b</td>
<td>20</td>
<td>1 a</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>735</td>
<td>695</td>
<td>94.6</td>
<td>8</td>
<td>2 b</td>
<td>25</td>
<td>1 b</td>
<td>12.5</td>
</tr>
<tr>
<td>Total</td>
<td>2910</td>
<td>2833</td>
<td>97.4</td>
<td>48</td>
<td>3 a, 6 b</td>
<td>18.8</td>
<td>3 a, 1 b</td>
<td>8.3</td>
</tr>
</tbody>
</table>

G is ground-truth; S is system detected; Scwdf is system classification without data fusion; Scudf is system classification using data fusion.

aMiss; bOver-counted.

![Fig. 8](https://example.com/fig8.png)

**Fig. 8** Speeds detected by this system and radar gun

a Test video A, correlation coefficient: 0.834
b Test video B, correlation coefficient: 0.927
c Test video C, correlation coefficient: 0.936
d Test video D, correlation coefficient: 0.895
5 Error investigation

The accuracy of video-based vehicle detection systems relates to several factors: accurate segmentation and extraction of vehicle contours, effective shadow removal, illumination changes, vehicle occlusions and so on. The LV with surface colour similar to the background colour cannot be segmented fully and extracted correctly and its length and area are hard to obtain accurately, which leads to misclassification. Some LVs are very large and occupy almost two lanes, which are counted twice. The system uses the FPVH as the feature to track the vehicle, so the tracking will be invalid when the vehicle head is blocked. For example, a car is following a truck and the car’s head may be blocked by the truck as long as the truck is enough large. The system recognises the truck and the following car as one LV under this condition, so the following car is missed. Data fusion indeed reduces the errors caused by vehicle occlusions especially the longitudinal occlusions, but it cannot solve the problem for long-time longitudinal occlusion. The shadow removal method used by this system is not self-adaptive to remove the shadows accurately, leading to some errors for classification and speed detection. That is why the speed detection result for test video A was not as good as others. The road surface is apparently wet and has frequent road reflection changes in the obviously raining weather, which leads to poor vehicle segmentation and serious errors. The system is effective in the drizzle weather but not useful in the obviously raining weather. The system will lose effectiveness under the extremely congested traffic conditions. The system also needs more considerations to improve its validity in the night. Therefore further researches are required to overcome these challenges and difficulties to improve the system robustness.

6 Conclusion

The system tracks every passing vehicle and records the type and speed of every passing vehicle. Finally, the traffic data of multiple vehicle types are collected. A colour image-based adaptive background subtraction is proposed to obtain more accurate vehicle objects, and a series of processes like shadow removal and setting road detection region are used to improve the system robustness. The cross-lane vehicles are detected and repeated counting for one vehicle is avoided, which improves the accuracy of vehicle counting. The system uses the space ratio of the blob and data fusion to reduce the classification errors caused by vehicle occlusions.

The accuracy of vehicle counting was 97.4% and the error of vehicle classification was 8.3%. The correlation coefficient of speeds detected by this system and radar gun was 0.898 and the mean absolute error of speed detection by this system was only 2.3 km/h. The traffic data of multiple vehicle types collected by this system are reliable and effective, which can be used in the fields of traffic
management and control, traffic simulation, pavement design and so on.

The further researches will pay attention to developing self-adaptive shadow removal methods, doing some online tests, detecting more traffic parameters, and comparing the system with the similar commercial products from ITS companies like Econolite, Iteris and Trafficon.

7 Acknowledgment

Financial supports provided by the National Natural Science Foundation of China (60804049) and the Special Fund for Basic Scientific Research of Central Colleges provided by Chang'an University (CHD2010JC022, CHD2012DK09) are acknowledged. The paper was also partly supported by the National High-Technology Research and Development (863) Programme of China (grant nos. 2011AA110404) in China.

8 References

4 Li, S., Yu, H., Zhang, J., Yue, K.: ‘Multi-type vehicles’ traffic data collection using video processing’. The Second Int. Conf. on Intelligent Control and Information Processing, 2011, pp. 271–276