Turk J Elec Eng & Comp Sci, Vol.19, No.4, 2011, © TÜBİTAK
doi:10.3906/elk-0910-266

Measuring traffic flow and classifying vehicle types: A surveillance video based approach

Erhan İNCE
Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10 TURKEY
e-mail: erhan.ince@emu.edu.tr

Received: 26.10.2009

Abstract

The paper presents a vehicle counting method based on invariant moments and shadow aware foreground masks. Estimation of the background and the segmentation of foreground regions can be done using either the Mixture of Gaussians model (MoG) or an improved version of the Group Based Histogram (GBH) technique. The work demonstrates that, even though the improved GBH method delivers performance just as good as MoG, considering computational efficiency, MoG is more suitable. Shadow aware binary masks for each frame are formed by using background subtraction and shadow removal in the Hue Saturation and Value (HSV) domain. To determine new vehicles in the subsequent frame (in addition to those in the current frame), invariant moments are used. For vehicles which are the same model and brand, color information and distance between center of mass and an imaginary reference line need to be considered. As for classification, the paper proposes a new method based on perspective projection of the scene geometry. The classification is grouped into three major tracks: bikes, saloon cars, and long vehicles. For each category, a lower and an upper bounding curve are developed to show the extent of their associated modality regions.

Key Words: *Group based histogram, mixture of Gaussians, cast shadow removal, convex hull fitting, classification, modality regions*

1. Introduction

In the past decades, to estimate the traffic flow or traffic density on a road, loop detectors or supersonic wave detectors were used by road engineers. Present-day traffic management systems on the other hand utilize image and video processing techniques to extract the same information from surveillance video of roads or junctions. In visual surveillance applications, detection by background subtraction is a common approach for differentiating moving objects from the static parts of the video frames. Background subtraction involves calculating a reference image, subtracting this reference from each new frame, and then applying a threshold to the result.

Through application of image and video processing techniques, a video based surveillance system (VBSS) can classify vehicles by their aspect ratios and report on traffic density and speed for each class and lane. In the literature several background modeling techniques have been described but only a few of these perform well in outdoor environments. The Mixture of Gaussians method (MoG) [1, 2] is very robust to multi-modal backgrounds and with some mean-variance normalization [3], can be strengthened to handle sudden changes in illumination. A more recent technique known as group based histogram (GBH) approach [4], is known to better cope with transient stops and with some extra effort could be made robust against multi-modal backgrounds (shaking leaves, swaying branches etc.) also.

Even though there are many techniques for background modeling, only a handful of techniques have been proposed for measuring traffic flow and density given recordings of surveillance videos. In [5], Takahashi introduced an image sensor for measuring traffic flow via vehicle tracking. The idea was to use an initial detection area in the center of the image and detect the front shape of vehicles to be used as templates. Then the image sensor would track front shapes of vehicles from previous to current frame by gray level pattern matching. Taniguchi proposed the use of a directional temporal plane transform (DTT) [6], which would transforms spatio-temporal images to 2-D data on a directional-temporal plane. Wakabayashi and Aoki in [7] introduced the use of stereo slit cameras where a top view slit camera with its slit direction parallel to lateral direction of road will give a pseudo two-dimensional top view image of vehicles. The stereo concept was used to gather height information in order to separate between vehicles and cast shadows. In [8], Tekalp and Salman make use of an adaptive bounding box size to detect and track vehicles according to their estimated distances from the camera given the scene-camera geometry.

In this paper, a new approach is proposed to track vehicles in consecutive frames. First the moments for all foreground (FG) objects in two consecutive frames are computed. Second, the moments of FG blobs in the latter frame are compared with those from the first. If after all comparisons the deviation in moments of a particular blob is larger than a previously determined error then this blob is assumed to be a new vehicle and the car count is incremented. For tracking, the boundaries of the current frame are used and only objects which do not violate these lines and fall entirely in the designated region are tested against ones from the previous frame. When two or three same-brand and/or model cars appear in two consecutive frames, firstly color and secondly distance information from center of mass of the blob to one of the reference lines need to be considered. Figures 1(a) and 1(b) depict the distance information both for a straight and meandering road. For the meandering road the reference point has been assumed to be the lower left corner of the image.

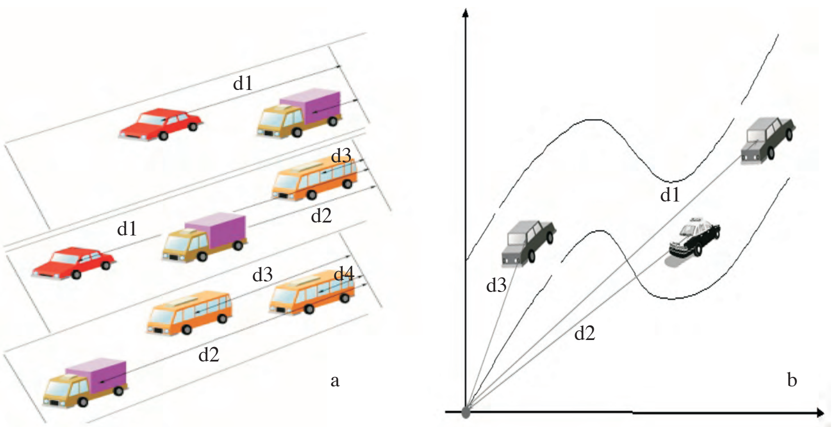

Figure 1. Identifying same brand cars by color or distance (a) traffic on a straight road; (b) traffic on a meandering road.

Because the proposed vehicle counting method only requires the computation of a set of distances between the detected object combinations in every other two frames, it does not require much storage and the possibility of real-time applicability is high. Also due to the shift, scale, and rotation invariance of moments, any orientation changes due to driver (lane changes and or minor deviations from the original path) will not affect the comparisons of the detected FG objects in consecutive frames. Similarly, classification step which is based on perspective projection of the scene geometry is also simple. Modality regions for three categories (bikes, saloon cars, and long vehicles) are developed offline and then the characteristic plot is used to classify objects. The processing is not complex, because classification is only dependent on to which region the detected object is closer.

Remainder of this paper is organized as follows. Section 2 presents a modified version of the group based histogram method for background modeling. It is shown that with some minor modifications the standard GBH can become as robust as the MoG model for non-static backgrounds. However analysis of the computational complexity shows that for real time applications MoG is still more appropriate. Section 3 underlines the need for shadow removal and gives examples based on the HSV color space. Section 4 proposes a new approach based on invariant moments to track and count the number of vehicles in a video sequence. Section 5 proposes a novel classification method based on modality regions and gives some insight as to how it can be applied. Section 6 provides a brief analysis on the complexity of the various tasks in terms of elapsed time and flops. Finally, section 7 ends the paper with some conclusions.

2. Constructing a robust background model

Though many background subtraction methods are listed in the literature foreground detecting specially for outdoor scenes is still a very challenging problem. The performance of the VBSS will vary based on lighting conditions, camera angle and height, weather conditions and camera vibrations due to wind and heavy vehicles. Ideally, background subtraction should detect real moving objects with high accuracy. However in practice the detection of cast shadows as foreground objects is very common since cast shadows also move at the same rate as the vehicles.

For outdoor applications the mixture of Gaussians (MoG) and the Group Based Histogram (GBH) technique are the preferred background estimation methods. The MoG algorithm can suppress the multi-modal backgrounds more effectively when compared with other methods [9]. On the other hand, the standard GBH method is known to better handle the transient stops but is not so robust against multi-modal backgrounds (shaking leaves, swaying branches etc). In this work we have adopted the standard GBH method described in [4] and have made some modifications in order to make the method more robust against the non-static backgrounds. The comparisons for the background estimation/foreground segmentation (BE/FS) of the two algorithms were based on "video7_long.avi", a synthetically generated sequence of VISOR [10]. Together with the test sequence the precision and recall metrics were employed.

To have a precise comparison between BE/FS algorithms video sequences with ground truths are necessary. These are video sequences that are created by first recording a scene without any foreground objects and then superimposing animated moving objects on the recorded background manually. Therefore, the exact location of the pixels related to foreground items is known (in other words the ground-truths of these sequences are available). A second advantage of using a test sequence with ground truth is that the superimposed objects would not contain shadows and hence the focus will be on the BE/FS performance only (don't have to deal with shadow removal at the same time).

Figures 2 and 3 show some sample frames and their corresponding ground truths for video7_long. Frame #545 depicts the scene without any foreground objects. Looking at the location of tree branches below the window in the top left corner of various frames we can say that at the day of recording some wind existed. Therefore to assess the robustness of the improved GBH against multi-modal backgrounds this video sequence is a good test sequence.

Having looked at distribution of scene intensities for a number of selected frames from the synthetically generated "video7_long," we noticed that there would be multiple peaks in the histogram of scenes with non-static backgrounds (refer to Figure 4(a)). The peak with the highest value (on the left) is representing the static parts of the background but the remaining two peaks (on the right side) are due to shaking leaves or swaying branches. The improved GBH model sets out to consider some of these smaller peaks. Those modes whose maximum values are greater than 50% of the maximum value for the tallest mode are also considered as constituents of the background model.

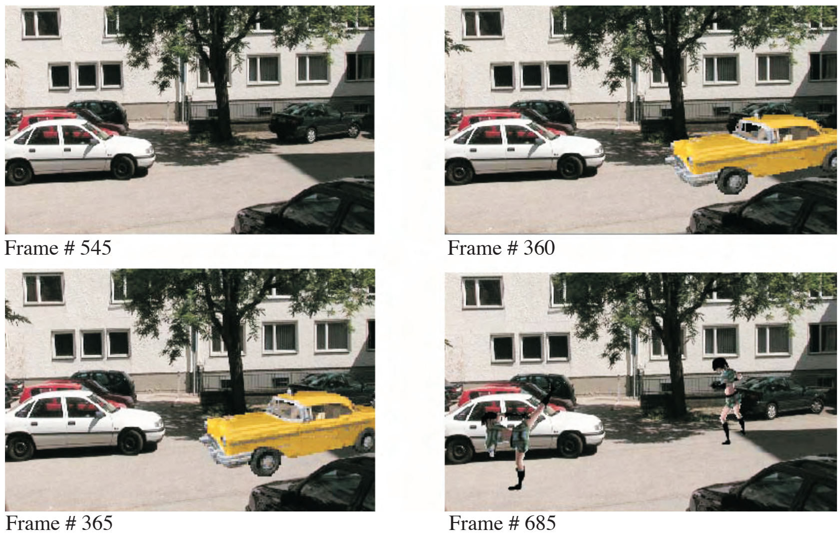

Figure 2. Sample frames from file "video7_long."

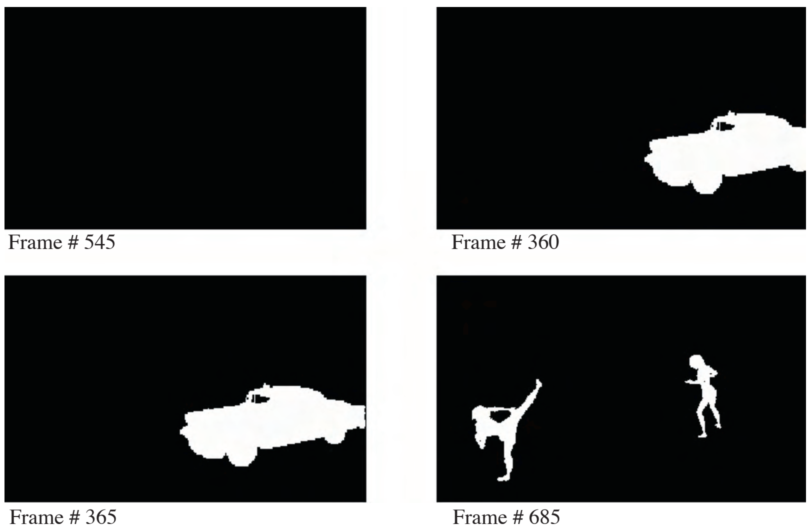

Figure 3. Ground truths for sample frames shown in figure 2.

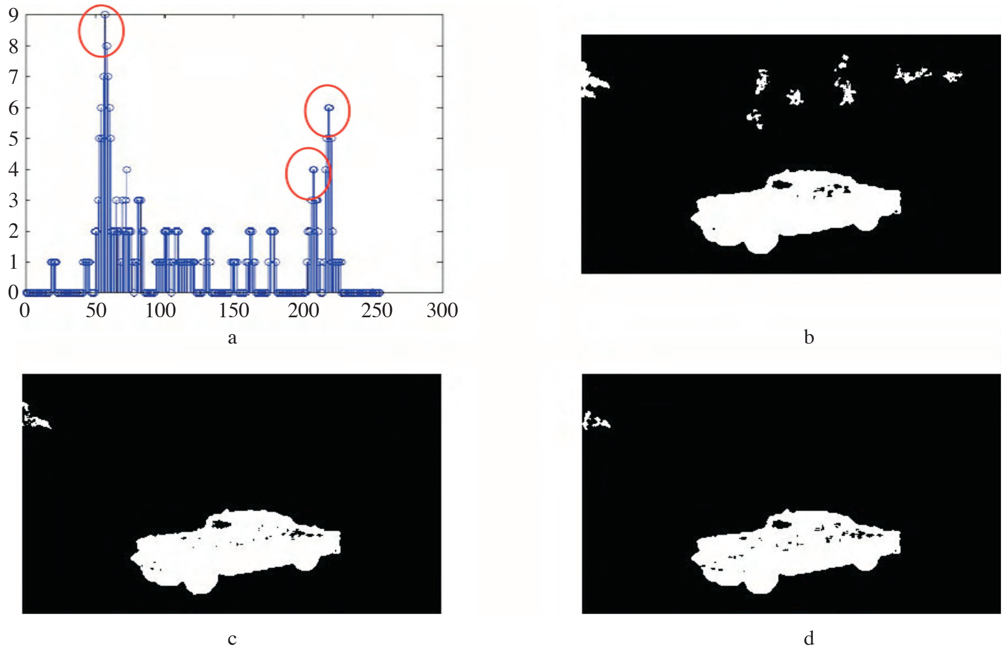

Figure 4. GBH-equipped image frames to combat multi-modal backgrounds, showing (a) image histogram with multiple peaks; (b) standard GBH; (c) an improved GBH; and (d) a MoG detected FG.

As can be seen directly from Figures 4(b) and 4(c), the improved GBH algorithm has become much more robust against multi-modal backgrounds and has managed to suppress the swaying leaves that were detected as foreground in its standard version. Figure 4(d) depicts the FG detected by the mixture of Gaussians model for the same frame. It is clear that the re-enforced GBH is performing just as well as the mixture of Gaussians model.

According to [11], comparison of algorithms trying to achieve the same task are possible either using *standalone evaluations* or by the application of *relative evaluation* methods. In this work the latter approach was used. To compare our achieved results with the ground truths two well known scales "recall" and "precision" were employed for each pixel. Recall is a measure of completeness and is defined as the number of correctly identified pixels (true positives) divided by the total number of pixels that actually belong to the foreground objects (pixels in ground truth). On the other hand, precision is defined as the ratio of correctly detected pixels in the region of interest to the number of all pixels in relevant detection regions:

$$R = \frac{TP}{TP+FN}$$
$$P = \frac{TP}{TP+FB} \tag{1}$$

During evaluations, all the frames belonging to "video7_long" were used and average recall and precision percentages were computed both for the MoG and the improved GBH. The results are shown in Table 1.

Table 1. Recall and Precision percentages for test video sequence "Video7_long.avi."

BE/FS Method	Recall (%)	Precision (%)
Improved GBH method	86.16	76.42
Mixture of Gaussians method	85.38	77.96

It is clear from the precision percentage that the improved GBH can segment the foreground almost as accurately as the MoG.

However, when average processing times were computed using again "video7_long" (384×240 frame resolution, 25 frames/sec) the results showed that improved GBH would require almost 3 times more processing time than the mixed-Gaussians method. This would further increase approximately by a factor of 3, since the resolution of the recordings taken in Famagusta city were 640×480. Hence considering real-time applicability MoG method will be a better choice.

3. Shadow detection

To get an accurate count on the number of vehicles passing through a particular part of the road two things are vital: first, a good background/foreground separation algorithm; and second, an efficient shadow detection and removal routine. For outdoor environments almost always there would be some cast shadows which are produced by partial or entire occlusion of direct light from a light source by the moving vehicles. Since cast shadows also move at the same rate as the vehicles, inevitably they are detected as part of the foreground. Long shadows could also connect two separate vehicles as if they were a single object. Therefore, the performance of the overall system may be seriously affected if the cast shadows are not detected and removed efficiently.

In this work in order to determine which pixels belong to the cast shadow region, the HSV color space has been exploited. The driving force behind this decision was the comparative study of shadow detection techniques carried out in [12] and [13]. Figure 5, shows examples of cast shadow detection and removal based on the HSV method. The test frames were all obtained from recordings of different roads in the Famagusta city.

Figure 6 also depicts an example of cast shadow removal applied to a sample frame of the Yeni İzmir junction in the Famagusta city. The only difference this time is that the convex hull fitting as explained in [14] has been applied. It is known that at times while trying to separate moving objects and cast shadows (which together constitute the foreground mask), some pixels belonging to the vehicle parts can be misclassified as shadow pixels. This can lead to partial erosion and holes in the FG mask. Before computing invariant moments based on these incomplete FG masks, it is possible to fit a convex hull to the exterior points for each vehicle and then create a new mask that includes all the points in the enclosed polygon.

4. Moments based flow measurement

Moment invariants are properties of connected regions in binary images that are invariant to translation, rotation and scale. They are useful because they define a simply calculated set of regional properties that can be used for shape classification and in part for recognition. As described in [15], the $(pq)^{th}$ central moment can be computed using the formula

$$M_{pq} \sum_{X} \sum_{y} (x - \hat{x})^p (y - \hat{y})^q f(x, y) \tag{2}$$

Translation invariance is due to the fact that the center of mass is subtracted during the computation of the values (as a part translated so does its center of mass). Also, the moments can be made scale invariant by dividing M_{pq} by $A^{1+(p+q)/2}$ where, A denotes the number of white pixels in the binary image. Note that, if a part doubles in size, then the value of M increases in proportion to its area and to the powers of the two moment terms. So, the denominator of the scale-invariant function increases by the same amount as M, and hence the ratio becomes scale invariant.

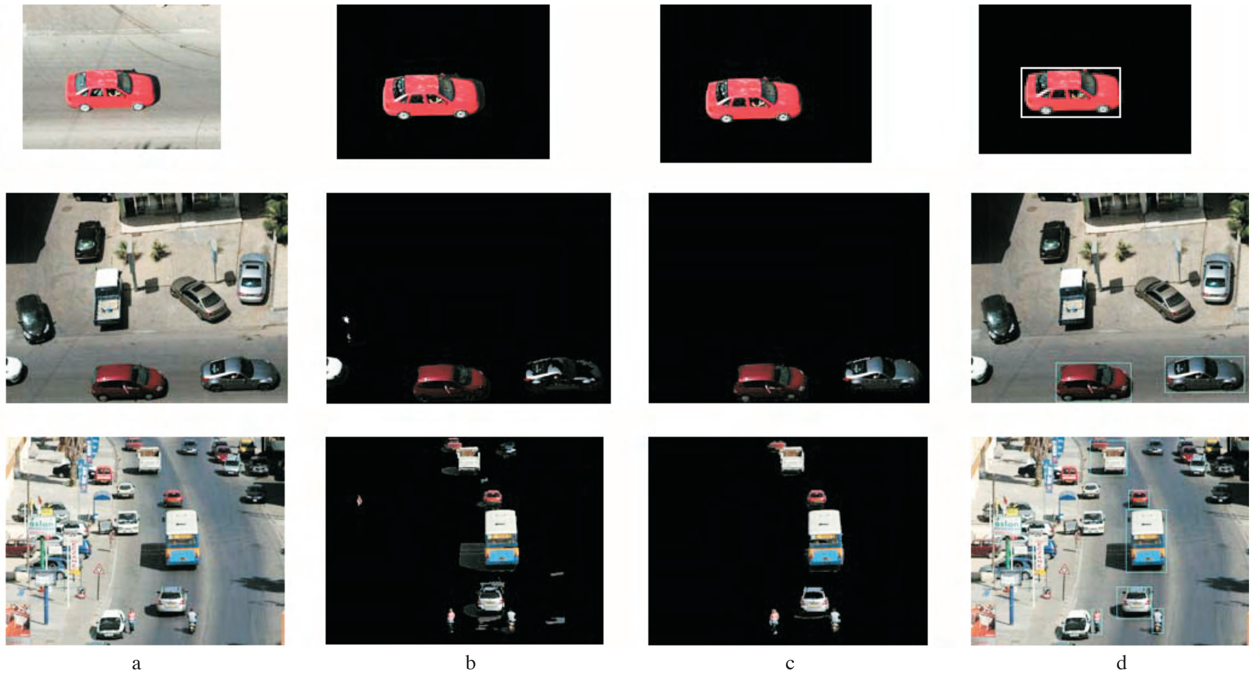

Figure 5. Shadow detection and removal using HSV domain statistics: (a) The original frame, (b) FG processed with cast shadow, (c) FG object(s), and (d) Localized vehicles.

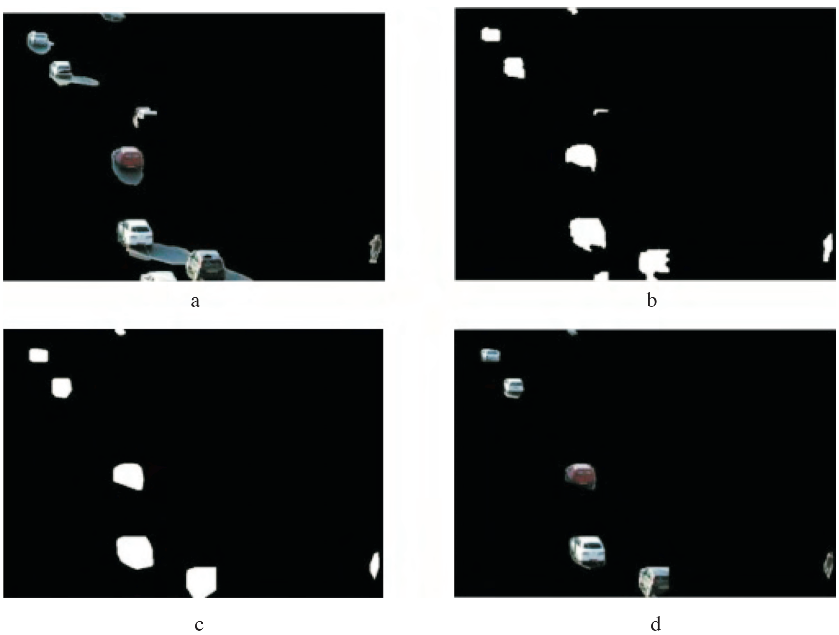

Figure 6. Cast shadow removal using the HSV color space and convex hull based mask: (a) FG with cast shadows; (b) FG mask after shadow removal (c) Convex hull based FG mask; (d) FG vehicles.

What follows, demonstrates the use of invariant moments for traffic flow measuring with the help of frames depicted in Figure 7. At the beginning, the moments for each detected moving object in frame #1 have been recorded. Then for the subsequent frames the Euclidean distance between the moments of each FG object and the vehicles detected in previous frame was computed. If the distance values for all pairs are above a particular threshold the mobile object would be counted as a new vehicle, otherwise the distance criteria mentioned in section 1 would be considered to see how far each object is from a reference line.

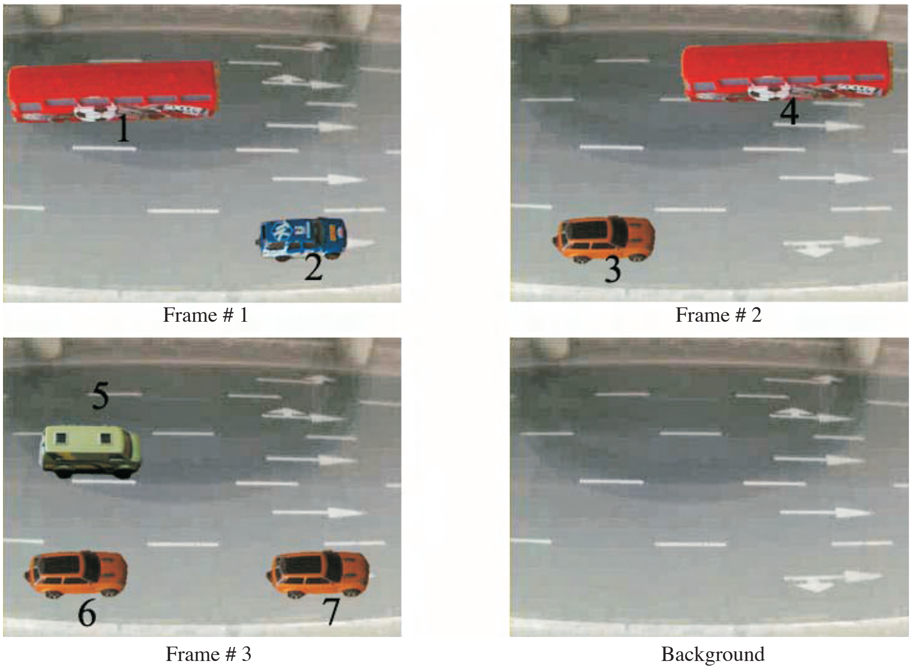

Figure 7. Counting vehicles in subsequent frames.

Table 2 provides four invariant moments based on the second and third central moments [15], of each FG object (1-2-3-4-5-6-7) detected. The table also contains the Euclidean distances between different pairs in consecutive frames.

Note that when the distance values are below unity we are either dealing with the same vehicle that has moved further away or with two cars that are the same brand and model. As can be observed from Table 2, the Euclidean distance for two different FG objects is significantly larger (approximately 5 times the reference value). Simulation results show that measuring traffic flow using the proposed approach favorably compares with other techniques in the literature. Also the proposed approach requires less computational complexity since it solely relies on moment based distance comparisons.

5. Modality region based classification

Following segmentation and connected-component analysis, the detected foreground blobs can be classified into pre-specified object categories including automobile, bus/truck and bike. In [16], Dedeoğlu and Çetin used the silhouettes of detected human objects and compared them against pre-labelled templates in an object silhouette database. Similarly Hu et al. [17], used feature vectors in terms of *centro-distances* to classify moving objects in mixed traffic. In [18], Chen proposed the use of temporal features such as *dispersedness*, *aspect ratio*, and

Table 2. Distances based on moments.

Vehicle No.	Invariant Moments Vector, ϕ
FRAME #1	
1	6.1248, 12.5643, 24.8620, 24.6346
2	5.7582, 12.5842, 21.5861, 21.2901
FRAME #2	
3 24.2589	6.1093, 13.1471, 24.0259,
4	6.1236, 12.5635, 25.2121, 24.2589
FRAME #3	
5	6.1140, 13.3066, 22.9993, 22.7363
6	6.1099, 13.1474, 24.0461, 24.2526
7	6.1098, 13.1265, 23.9712, 24.2505
Euclidean Distances	
d_{31}= 3.2158	d_{32} = 10.7257
d_{41} = 0.7181	d_{42} = 12.9139
d_{53} = 6.7586	d_{54} = 8.9830
d_{63} = 0.2844	d_{64} = 3.7336
d_{73} = 0.4480	d_{74} = 3.9775

area ratio. These temporal features greatly reduce the problem of changeable features extracted from a moving object.

We propose in this paper a new classification approach based on perspective projection of the scene geometry. As can be seen from Figure 8(a), for a particular vehicle on the road (world plane) the area occupied over time (in different frames) will change in relation to distance covered, the type of lens, and its focal length. By capturing a series of images at varying locations (x, y) for vehicles selected from different categories one can easily develop a modality curve (curves related to structure) for each category. Each curve would depict the projected area versus distance for a category as shown in Figure 8(b).

Each category is associated with an upper bound and a lower bound since one has to assume some lower and upper limits for widths and heights of vehicles in the same category. For example, a mini car with a width and length of 1.3 m $\times$ 2 m can constitute the minimum value set for automobile; and a Honda Civic which is 2.3 m $\times$ 3.3 m form the maximum values. After developing two curves per category, a new vehicle can be categorized based on its area versus distance value and its Euclidean distance from the different curves.

It is also possible to develop an analytical formula for estimating the area occupied on the image plane by a particular vehicle. Assume that a pinhole camera with a convex lens is used to observe an approaching or departing vehicle and that the location of the camera (height from ground) is known (refer to Figure 9). Through perspective projection of the scene geometry using the lens equation (3) and the magnification factor (4), we can derive a mathematical formula as follows:

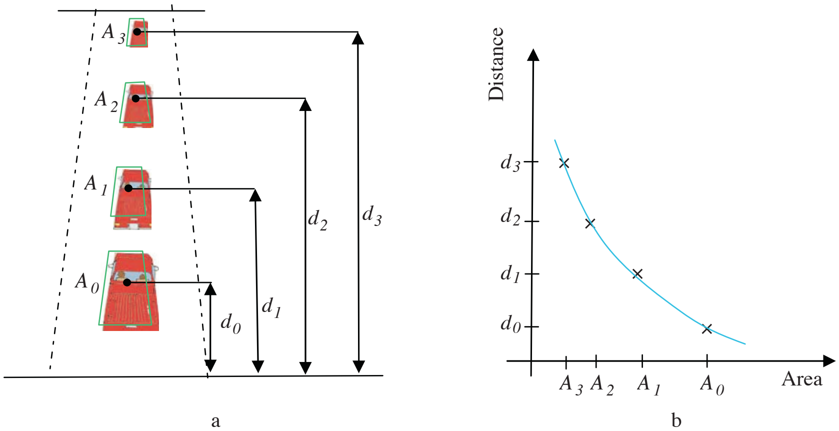

Figure 8. Area occupancy on Image Plane.

$$\frac{1}{f} = \frac{1}{d_{obj}} + \frac{1}{d_{im}} \tag{3}$$

$$M = \frac{h_{im}}{h_{obj}} = \frac{|d_{im}|}{d_{obj}}. \tag{4}$$

As seen from Figure 9, the area of the vehicle in the world plane would be $l_{obj} \times 2w_{obj}$ where, $w_{obj} = \frac{|d_{obj}-f|}{f} w_{im}$ and $l_{obj} = \frac{|d_{obj}-f|}{f} l_{im}$ from equations (3) and (4). By these and noting the fact that the area in the image plane would be $l_{im} \times 2 \cdot w_{im}$ and $d_{obj}^2 = h_c^2 + d_g^2$ we can obtain the equation

$$A_{im} = \frac{f^2}{h_c^2 + d_g^2 + f^2 - 2f\sqrt{h_c^2 + d_g^2}} A_{veh}. \tag{5}$$

Assuming a 35mm camera positioned at a height of 12m and using equation (5) a set of lower and upper bounds have been generated (see Figure 10) for bicyclists, saloon cars, and long vehicles using the min and max vehicle dimensions provided in Table 3. Prior to recording of the video sequences, some road markers (traffic cones) were placed on the side of the road at pre-selected distances. Then the video was taken and background model developed. Finally, selecting a frame in which the test vehicle was across from one of these road markers FG separation and counting of the white pixels corresponding to the test vehicle would give us the area occupied in the image plane. Figure 10 depicts three green markers for three test vehicles, one from each category, at different ground distances. The triangular marker is for a bike, the diamond is for a saloon car of dimensions 1.35 m × 2.6 m, and the pentagram represents a long vehicle of dimensions 3.1 m ×5.8 m at d_g of 38 m. Clearly a vehicle whose dimensions are not far from the dimensions of its category could easily be identified by this method.

A final consideration here is that, in daily practice no markers would be present on the side of the road. Therefore the ground distances of the vehicles would have to be obtained from the surveillance video. This is possible by following the same approach proposed by Bowen et al. in [19]. Using the distances between the test vehicle and a reference plane, together with areas occupied in the FG image by that vehicle a generalized curve could be fitted. Then this curve could be used to identify the distance of a vehicle in a particular frame based on the area the test vehicle occupies in pixels.

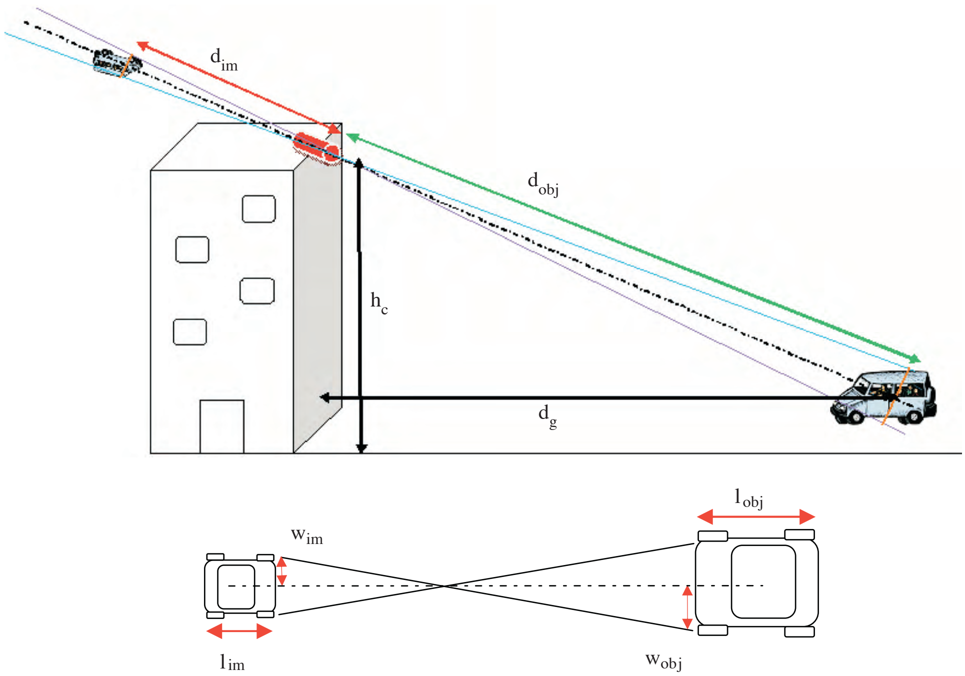

Figure 9. Perspective projection of the scene geometry.

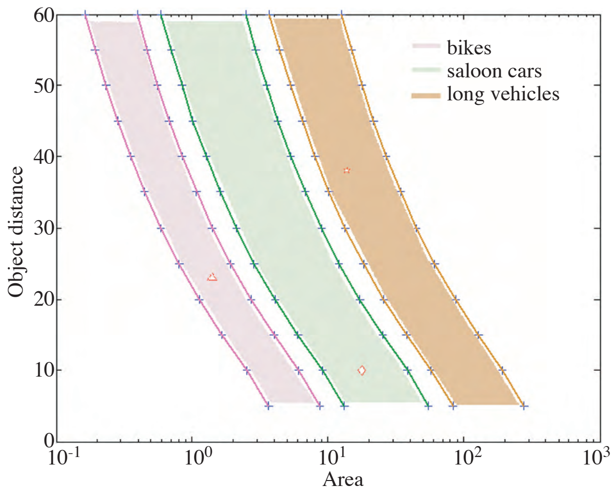

Figure 10. Modality Regions with lower and upper bounds for each category.

Table 3. Minimum and maximum vehicle dimensions in each category.

	Bikes	Saloon Cars	Long Vehicles
	min ∼ max	min ∼ max	min ∼ max
Width (m)	0.5 ∼ 0.6	1.2 ∼ 2.3	3.0 ∼ 3.2
Length (m)	1.0 ∼ 2.0	1.5 ∼ 3.3	3.8 ∼ 12.0

6. Computational efficiency analysis

Two measures of the efficiency of an algorithm are known to be: the elapsed time, and the number of floating point operations (flops) required by the mathematical operations. On time-sharing machines, elapsed time may not always be a reliable measure of the efficiency, since the rate of execution depends on how busy the computer is at the time. In this study, the computational complexity for each algorithm was shown both in term of elapsed time and the average number of flops required to process a single frame from the test video.

Because *flops* function has been discontinued after version 6 of MATLAB, in this study the MATPAPI303 API from Innovative Computing Laboratories [20], was installed to compute the number of floating point operations required. The efficiency tests were carried out using a HP lab-top computer with an Intel CPU operating at 1.83 GHz and a RAM of 1 GB. The average time taken per frame and the number of floating point operations required to complete each task is shown in Table 4. Clearly, the most computationally complex task is still the BE/FS and the total average time for all tasks to be completed is $\sim$11.5 seconds.

Assuming that the video camera is observing a 50 meters long section of a road and the vehicles are travelling around 25 km/hr (6.94 m/s), this means that a vehicle will completely move from one side of the frame to the other in 7.2 seconds. Hence processing with MATLAB still will not be in real time. However if the same code is implemented in a high level language, the time requirement is expected to reduce by a factor of 10 and in this case real-time processing will become possible.

Table 4. Average number of Flops per frame of processing.

	Average time elapsed, in seconds per frame	Average number of flops per frame
Background estimation and FG Segmentation	10.16	165304680
Shadow detection and removal using HSV	0.79	82453340
Moment computation(s)	0.57	24458493

Further computational efficiency can also be gained by the application of multithreading background subtraction technique as described in [21]. It has been stated in [21] that, speed gains by a factor of 2.5 – 4 would become possible through the use of multiple threads.

7. Conclusions

Prior to obtaining traffic parameters, a successful separation of foreground parts from frames of the surveillance video is most crucial. In this paper for the segmentation of foregrounds from the scene an improved version of the GBH background estimation method has been proposed. By considering all other peaks in the histogram whose maximum values are greater than 50% of the main mode, it was shown that the new GBH can perform just as good as the mixture of Gaussians model when the scene contains non-static pixels. During the comparisons a synthetically generated video with ground truths for each frame was used together with precision and recall scales. When average processing times were compared it became evident that the improved GBH would require almost 3 times more processing time when compared to the mixture of Gaussians method.

The present paper has also presented a vehicle counting method based on invariant moments and shadow aware foreground masks. Initial results using the proposed approach and some custom videos taken at different locations in Famagusta city indicate that this approach is computationally efficient. With MATLAB processing, computation of moments for one frame on the average took a little more than half a second. Since the proposed method needs to compare two consecutive frames the full processing on the average will require anywhere between 1–2 seconds. However, since the most time consuming part is still the BE/FS, a full real time implementation would not become possible without using either a high level language or by moving towards more efficient customized solutions such as ASICs.

Finally, also proposed was a new framework for classifying vehicles in traffic based on their modality curves. For each category a lower and an upper bound is generated. It is demonstrated that a new test vehicle can be categorized based on its area versus distance values and its Euclidean distance from the different curves. The present paper also derives a mathematical formula for calculating the area occupied on the image plane by a particular vehicle on ground.

Even though the primary results on classification are promising more work has to be done to obtain the intrinsic parameters of the camera following the work presented by Bowen in [19].

Acknowledgements

This work was supported by the Ministry of Education and Culture of T.R.N.C. under a grant with contract number BAP-290/115. Special thanks go to Prof. Dr. Runyi Yu for fruitful discussions and insightful comments.

References

[1] C. Stauffer and WEL. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *IEEE Conf. on Computer Vision and Pattern Recognition,* vol. 2, pp. 246–252, 1999.

[2] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *European Workshop on Advanced Video Based Surveillance Systems*, AVBS2001, pp. 149–158, Sep 2001.

[3] P. Pujol, D. Macho and C. Nadeu, "On real-time mean and variance normalization of speech recognition features, *ICASSP*, pp. 773–776, 2006.

[4] K.Song and J.Tai, "Real-Time Background Estimation of Traffic Imagery Using Group-Based Histogram," *Journal of Information Science and Engineering*, vol. 24, pp. 411–423, 2008.

[5] K. Takahashi, T. Kitamura, M. Takatoo, Y. Kobayashi and Y. Satoh, "Traffic Flow Measuring System by Image Processing," *IAPR Workshop on Machine Vision Applications*, pp. 245–248, Nov 1996.

[6] H.Taniguchi, T. Nakamura and H. Furusawa, "Methods of traffic flow measurement using spatio-temporal image," *Int. Conf. on Image Processing*, vol. 4, pp. 16–20, Oct 1999.

[7] Y. Wakabayashi and M. Aoki, "Traffic Flow Measurement Using Stereo Slit Camera," IEEE *Conf. on Intelligent Transportation Systems*, pp. 727–732, Sept. 2005.

[8] E. Baş, A. M. Tekalp and S. Salman, "Automatic Vehicle Counting from Video for Traffic Flow Analysis," *IEEE Intelligent Vehicles Symposium*, pp. 392–397, June 2007.

[9] N. Seifnaraghi, "A Comparative Study of Background Estimation Algorithms," *MS Thesis*, Eastern Mediterranean University, Sep 2009.

[10] Video surveillance online repository: http://www.openvisor.org/.

[11] V. Mezaris, I. Kompatsiaris and M.G. Strintzis, "Still Image Objective Segmentation Evaluation using Ground Truth," 5^{th} *COST 276 Workshop*, pp. 9–14, 2003.

[12] A. Prati, I. Mikic, C. Grana and M. M. Trivedi, "Shadow Detection Algorithms for Traffic Flow Analysis: Comparative Study," *IEEE Int. Conf. on Intelligent Transportation Systems*, pp. 340–345, Aug 2001.

[13] S. G. Ebrahimi, "Shadow Aware Object Detection and Vehicle Identification via License Plate Recognition," *MS Thesis*, Eastern Mediterranean University, Sep. 2009.

[14] N. Seifnaraghi, S. G. Ebrahimi and E. A. İnce, "Novel Traffic Lights Signaling Technique Based on Lane Occupancy Rates," *Int. Symp. on Computer and Information Sciences*, pp. 592–596, Sep 2009.

[15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall, 2002, pp.672–675.

[16] Y. Dedeoğlu, B. U. Töreyin, U. Güdükbay, and A. E. Çetin, "Silhouette-Based Method for Object Classification and Human Action Recognition in Video," *HCI/ECCV 2006*, LNCS 3979, pp. 64–77, 2006.

[17] H. Hu, A. Hu, A. Li and D. Wang, "Research on Recognition and Classification of Moving Objects in Mixed Traffic Based on Video Detection," *Transportation Research Board Annual Meeting* , Paper #08-1437, 2008.

[18] T-H. Chen, Y-F. Lin and T-Y. Chen, "Intelligent Vehicle Counting Method Based on Blob Analysis in Traffic Surveillance," *Second International Conference on Innovative Computing, Information and Control*, pp. 238–242, 2007.

[19] R.M. Bowen, E. Çınar and F. Şahin, "System of Systems Approach to a Human Tracking Problem with Mobile Robots using a Single Security Camera," 5^{th} *Int. Conf. on Soft Computing, Computing with Words and Perception in System Analysis, Decision and Control*, Sep. 2009.

[20] J. Dongarra, K. Landon, S. Moore, P. Mucci, D. Terpstra, J. Thomas, H. You and M. Zhou, "PAPI 3 for MATLAB," Innovative Computing Laboratory.

[21] S. C. Lee and R. Nevatia, "Speed performance improvement vehicle blob tracking system," Multimodal Technologies for Perception of Humans: International Evaluation Workshops CLEAR 2007 and RT 2007, pp. 197–202, 2008.