Real-time Road Traffic Density Estimation using Block Variance

Kratika Garg Siew-Kei Lam Thambipillai Srikanthan
Nanyang Technological University, Singapore
kratika001@e.,assklam@,astsrikan@ntu.edu.sg

Vedika Agarwal
Birla Institute of Technology and Science, Pilani, India
vedikaagarwal10@gmail.com

Abstract

The increasing demand for urban mobility calls for a robust real-time traffic monitoring system. In this paper we present a vision-based approach for road traffic density estimation which forms the fundamental building block of traffic monitoring systems. Existing techniques based on vehicle counting and tracking suffer from low accuracy due to sensitivity to illumination changes, occlusions, congestions etc. In addition, existing holistic-based methods cannot be implemented in real-time due to high computational complexity. In this paper we propose a block based holistic approach to estimate traffic density which does not rely on pixel based analysis, therefore significantly reducing the computational cost. The proposed method employs variance as a means for detecting the occupancy of vehicles on pre-defined blocks and incorporates a shadow elimination scheme to prevent false positives. In order to take into account varying illumination conditions, a low-complexity scheme for continuous background update is employed. Empirical evaluations on publicly available datasets demonstrate that the proposed method can achieve real-time performance and has comparable accuracy with existing high complexity holistic methods.

1. Introduction

It has been projected that the number of vehicles in the industrialized world will double to 1 billion, while a 12-fold increase is expected in the developing world by 2050 [17]. With this increase in the number of vehicles, traffic congestion is bound to come up as a serious issue. Over the last few years, intelligent transportation systems (ITS) has become increasingly popular for dealing with the problem of traffic congestion. Road traffic density estimation is the basic step used in ITS for road planning, intelligent road routing, road traffic control, network traffic scheduling, routing and dissemination [20].

Conventionally, inductive loop detectors, wireless vehicle sensors and traffic surveillance cameras have been used for road traffic density estimation. Among these, vision-based methods pose a greater advantage as they incur low installation costs, little traffic disruption during maintenance and provide more coverage [16]. However, existing traffic monitoring systems which use video feeds from overhead stationary cameras monitoring a road segment suffer from delayed traffic updates and slow responsiveness to emergency situations. This is contributed by the fact that these systems rely on transmission of videos/images from the surveillance cameras to a central system that are then analyzed manually.

Several techniques have been proposed in the literature to automate this process. The existing approaches for traffic surveillance can be widely divided into three categories - Vehicle Counting, Vehicle Tracking and Holistic methods.

Vehicle Counting methods rely on moving object segmentation for traffic analysis. The techniques for moving object segmentation can be divided into four main categories - Frame differencing, Background Subtraction, Object based methods and Motion based methods. Frame differencing methods are easy to implement but they cannot deal with noise, abrupt illumination changes and periodic changes in the background [26][7]. Background Subtraction techniques are widely used in the literature due to their robustness in dealing with illumination changes, but the sophisticated background subtraction methods *e.g.* Hidden Markov Models and Neural Networks [19] which can deal with various environmental variations incur high computational costs [24]. Object based methods [28] try to identify complete objects using 3D models and Motion based methods [25] use optical flow to detect the moving objects. Both these methods have high complexity in terms of computations making them infeasible for real-time applications on low-cost platforms.

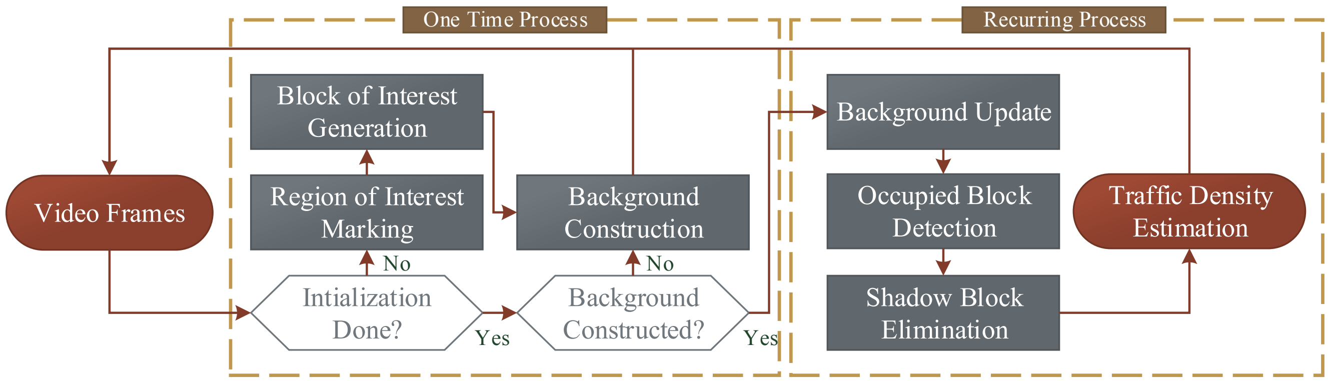


Figure 1: Overview of the proposed approach.

Moving shadow detection and removal is another crucial step in vehicle counting methods. Different methods based on colour [1], texture [15], physical properties [18] and geometry [10] have been proposed in the past 20 years. Although, texture based methods have been identified as the most accurate, their computational complexity is higher than all other methods proposed in the literature [23]. Thus moving shadow detection methods face a trade off between robustness and computational complexity. In addition to their sensitivity to illumination changes, and challenges in dealing with moving shadows, most vehicle counting methods tend to fail during traffic congestions as they group several vehicles together.

Tracking based methods [5][6] combine vehicle segmentation and tracking to calculate the velocity of the moving vehicles to estimate the traffic flow. In addition to the issues related to vehicle segmentation, these techniques also suffer from poor performance due to vehicle correspondence and occlusions.

More recently several holistic approaches have been proposed in the literature for classification of traffic videos. These techniques deal with whole image globally thereby avoiding segmentation of each moving object. Chan *et al.* first modeled the traffic video classification problem as a dynamic texture classification problem [4]. After that [13] and [9] also used the dynamic texture model based on Spatio-temporal Gabor Filters and 3D Spatio-temporal orientation energy respectively for classifying traffic videos. Classification of traffic videos using symbolic features is proposed in [8]. In [2], a combination of macroscopic (holistic) and microscopic (object-based) has been used to classify traffic videos. All these methods achieve a very high accuracy in video classification. However, the computational load of fitting their models for the classification process is very high.

Overall, vehicle counting and tracking methods which could be used real time are more sensitive to environment conditions and tend to fail during congestions, while holistic approaches which are invariant to environmental conditions would require specialized hardware for real-time implementation. A complete review of existing vision-based techniques for traffic surveillance systems can be seen in [16] and [3].

In this paper we present a novel technique for road traffic density estimation which overcomes the issues faced by existing techniques. A block based approach has been used to estimate lane-wise road traffic density. Each lane is divided into several blocks, and the percentage occupancy of a lane is calculated by detecting the blocks which are occupied by vehicles. The overall percentage occupancy gives a quantitative estimate of the traffic density on the road segment. Our proposed method is closer to a holistic approach, as each vehicle is not being localized, while the percentage occupancy of the entire image is being calculated to estimate traffic intensity.

The main contributions of this paper can be summarized as follows: (1) A camera perspective invariant technique for dividing lanes into blocks has been proposed. (2) A block based background construction and update method has been proposed which only uses the intensity variance of blocks, thus, has very low computational complexity. (3) A vehicle block detection technique has been used which can deal with illumination changes and can robustly differentiate between vehicles and shadows. Extensive evaluations on publicly available datasets with challenging conditions - illumination changes, moving shadows, different camera perspectives have been done which demonstrate the robustness of the proposed approach. It shows that, the proposed technique has comparable accuracy to state-of-the-art methods and is suitable for real-time implementation.

The remaining paper is organized as follows. Section 2 explains the proposed approach in detail. In Section 3, the proposed technique has been evaluated and compared with other state-of-the-art methods. Finally in Section 4, we draw conclusions.

2. Proposed Approach

A simple and effective way to estimate traffic density is to calculate the amount of road surface that is occupied. In this paper, we present a block based processing approach to calculate the percentage of the occupied road segment. A two step method is used. First is a one time process which involves Region of Interest (ROI) marking, Block of Interest (BOI) generation and background construction. The second step is a recurring process which involves background update, occupied block detection, shadow block elimination and traffic density estimation. An overview of the proposed approach can be seen in Fig. 1.

2.1. One Time Process

2.1.1 Region of Interest (ROI)

Region of Interest (ROI) can be static/moving depending on the application. For lane-wise traffic density estimation, we have a static ROI i.e. lanes since the camera is stationary. For our technique, we manually mark the lane boundaries using two lines, for each lane, to get the ROI. It can be visualized in Fig. 2a. This is a one-time process which can be performed at initialization or can be automated using a lane detection algorithm. This context-aware decision to mark ROI significantly reduces the amount of pixels that have to be processed for each frame in a video.

2.1.2 Block of Interest Generation

Once the lanes are marked, each ROI is further divided into blocks of interest or BOIs. In Fig. 2b the yellow blocks represent the blocks of interest in each lane. For our proposed method, only these BOIs are used for further processing.

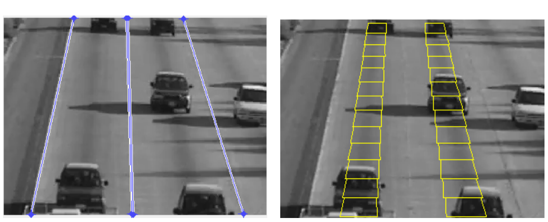
A camera perspective invariant technique was developed to divide each ROI into BOIs, which is described as follows. In this paper, we estimate traffic density by calculating the percentage of occupied blocks on the road. To get a correct estimate of percentage occupancy, each block should be smaller than the length of the smallest vehicle. Since, if a block length is larger, it would lead to over estimation of the percentage occupancy. It was observed that there is a relation between the width of a lane and length of a small vehicle. The length of a small vehicle and width of a lane remains in a set range in the 3D world, their ratio in an image is also expected to lie in a fixed range. In order to test this hypothesis, this value was calculated for all the datasets used and it was found that the ratio lies in a small range. The values differ slightly mainly due to camera perspectives and varying lane widths in different countries. For a lane width L_w and vehicle length V_l in pixels, the ratio can be defined as follows:

$$\lambda = L_w/V_l \tag{1}$$

We used $\lambda = 1.8$ to automate BOI Generation technique.

From the ROI Marking process, the lane width can be easily calculated, at each point in the lane. To generate BOI, starting from the bottom of the lane, L_w and corresponding $V_l = L_w/\lambda$ is calculated. This gives us a block with width = L_w and length = V_l. Since for our technique, we want the block size to be smaller than the vehicle size, this block is further divided into three equal horizontal blocks. Each generated block is further divided into three vertical divisions and central vertical division is defined as a BOI.

It can be visualized in Fig. 2b that the length of each BOI is approximately equal to 1/3rd of a small vehicle. We limit the number of BOIs per lane to 15 in order to ensure good visibility of vehicles.



(a) (b)

Figure 2: (a) Region of Interest (b) Block of Interest

2.1.3 Background Construction

The proposed background construction method is based on the variance of the pixel intensities in a BOI. When no vehicle passes through a BOI, its variance is expected to be the same across frames. This property holds even when there is an illumination change, since the intensities of all the pixels in the block would change together causing the variance to remain the same. Thus, the variance of the variance values across these frames for a BOI is expected to be low when no vehicle passes through it. We use this property to construct our background.

For each BOI, a circular buffer $buff_{BOI}$ is constructed which stores the variance values of N most recent frames. At the start of background construction, once the buffer is full, the variance of the stored values is calculated for each BOI i.e. VoV (Variance of Variances) which can be defined as:

$$VoV = Var(buff_{BOI}(:)) \tag{2}$$

If $VoV < T_{VoV}$, the pixel intensities of BOI from the current input frame are copied to the background image. The whole process is repeated until background is constructed for all BOI. For our proposed method, extensive simulations were conducted to generate the optimum value of N and T_{VoV}. N was set to 4 and T_{VoV} was set to 100. Our extensive simulations revealed that increasing the no.

of frames beyond 4 led to increase in the time taken for background construction, while the constructed background was the same. On the other hand, reducing the number of frames led to the deterioration of the background. For T_{VoV}, it was observed that that the difference between the VoV values with/without vehicle presence in a BOI for the past four frames was very high ($> 10^3$). To ensure robustness in background construction a low threshold of 100 was selected. After several experiments it was revealed that slightly decreasing/increasing the threshold did not lead to any change in the overall performance of the proposed approach. It only led to a slight increase/decrease in the time for background construction.

2.2. Recurring Process

2.2.1 Background Update

The background is updated at every frame to adapt to illumination changes and formation/fading of static shadows on the road. The background update procedure used is same as the above discussed background construction method. For each frame, $buff_{BOI}$ is updated and VoV is calculated. When $VoV < T_{VoV}$, the background is updated.

2.2.2 Occupied Block Detection

Once the background is constructed, the blocks occupied by vehicles have to be detected to estimate traffic density. The technique proposed to detect the occupied blocks is based on the observation that if a vehicle passes a block the intensity variance of that block would differ significantly from that of the background. The normalized variance difference w.r.t to the background for a BOI can be defined as

$$\Delta V = \frac{abs(Var^M_{BOI} - Var^I_{BOI})}{max(Var^M_{BOI}, Var^I_{BOI})} \tag{3}$$

where the subscripts M and I signify background and new video frame respectively.

Although, this parameter fails in cases when the texture of a vehicle part is similar to that of the background. Even when there the texture is similar, there would be an intensity difference between the background and the foreground pixels. Thus, in order to cope with such failures we calculate another parameter i.e. the percentage of foreground pixels in the BOI. Since the width of our BOIs are smaller than cars, the percentage of foreground pixels is expected to be high for vehicles. The foreground pixels are generated from a thresholded difference image. This parameter can be defined as

$$\%FG = \frac{\text{Foreground Pixels in BOI}}{\text{Total Pixels in BOI}} \tag{4}$$

It should be noted that foreground pixels have not been solely used for detecting occupied blocks because they

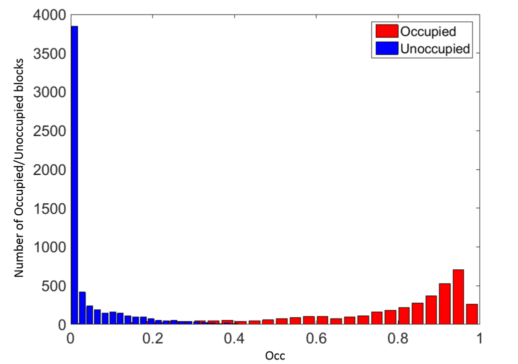


Figure 3: Occ Histogram for Occupied/Unoccupied Blocks

are more susceptible to background noise, illumination changes, shadows etc. which adds to a large number of false positives.

Finally, we used the geometric mean of the two parameters to classify the blocks, which is defined as follows:

$$Occ = \frac{2 * \Delta V * \%FG}{\Delta V + \%FG} \tag{5}$$

To analyze the effectiveness of Occ a statistical analysis was performed using 500 training images. For each image, each BOI was annotated as an occupied/unoccupied block. It should be noted that blocks containing cast shadows were also as annotated occupied blocks due to their similarities to vehicle occupied blocks. Finally two histograms were plotted for occupied and unoccupied blocks respectively. Fig. 3 shows a clear distinction between the histograms for occupied and unoccupied blocks.

Using Occ each block was classified into occupied and unoccupied blocks i.e. OB and UOB respectively.

$$\text{BOI} = \begin{cases} \text{OB} & Occ \geq T_O \\ \text{UOB} & Occ < T_O \end{cases} \tag{6}$$

An optimum threshold T_O was determined from Fig. 3 and was set to 0.3.

2.2.3 Shadow Block Elimination

In addition to detecting vehicle blocks, the occupied block detection method also detects moving shadow blocks as they have a variance difference comparable to vehicle blocks. Thus, the occupied blocks (OB) includes vehicle occupied block (VOB) as well as shadow occupied blocks (SOB). In this section we employ a shadow block elimination technique to get rid of SOB.

When a shadow falls on a road, the texture of the road remains preserved. Several shadow elimination techniques have used this property to eliminate shadows. Normalized Cross Correlation (NCC) is one of the techniques used to calculate the similarity between the background and shadow pixels [14]. In [14], NCC has been defined as Eq. 7 where I and M represent video frame and background respectively. A (2N+1)*(2N+1) neighborhood centered at pixel (i,j) is employed to calculate the NCC value. For our technique, N has been set to 1.

$$NCC(i,j) = ER(i,j)/\sqrt{E_M(i,j)E_I(i,j)} \tag{7}$$

where,

$$ER(i,j) = \sum_n \sum_m M(i+n,j+m)I(i+n,j+m)$$

$$E_M(i,j) = \sum_n \sum_m M(i+n,j+m)^2$$

$$E_I(i,j) = \sum_n \sum_m I(i+n,j+m)^2$$

$$-N \leq n \leq N; -N \leq m \leq N$$

This is a computational expensive calculation which includes several multiplications and square root calculations. In order to reduce the complexity of NCC calculation, we have taken the logarithm of Eq. 7. The modified equation is as follows:

$$\begin{aligned} &\log(NCC(i,j)) \qquad (8)\\ &= \log(ER(i,j)) - \frac{1}{2}(\log(E_M(i,j)) + \log(E_I(i,j))) \end{aligned}$$

Log calculations can be made compute-efficient for embedded devices using look up tables, hence this simple technique reduces a lot of computations. For our approach, only the foreground segmented pixels in BOI are used to detect shadow blocks, which limits the number of pixels for which NCC is calculated. The pixel (i,j) is pre-classified as shadow pixel if,

$$(log(NCC(i,j)) > T_{ncc}) \text{ and } (E_I(i,j) < E_M(i,j)) \tag{9}$$

For our proposed approach $T_{ncc} = log(0.90)$ as used in [14]. Although NCC serves a great measure to detect shadow pixels, it also wrongly classifies dark vehicle pixels as shadows. In order to prevent misclassification of dark objects, existing methods based on NCC for shadow elimination, combine it with a refinement stage. In this stage, the intensity ratio between foreground and background pixels is used to differentiate shadow pixels and dark object pixels. We have used the modified intensity ratio due to its ability to deal with shadows as well as reflections on the road [11]. For a pre-classfied pixel (i,j), the ratio R can be defined as

$$R(i,j) = (I(i,j) - M(i,j))/(I(i,j) + M(i,j)) \tag{10}$$

If $R(i,j) > T_R$ the pixel is classified as a shadow. It has been highlighted in [11] that R(i,j) for pixels corresponding to dark objects or shadow regions near objects lie in the range [-0.7,-0.1], while cast shadow pixels lie in the range [-0.5, -0.4]. Hence T_R was set to -0.5. Once all the segmented pixels in the BOI are classified as shadow/non-shadow pixels, the detected occupied blocks are classified into shadow and vehicle blocks.When more than 90% of the pixels in the BOI are classified as shadow pixels, that block is classified as SOB. This high threshold ensures that BOIs which are covered by both vehicles and shadows are classified correctly.

$$\text{OB} = \begin{cases} \text{SOB} & SB = 1 \\ \text{VOB} & SB = 0 \end{cases} \tag{11}$$

where,

$$SB = \left(\frac{\text{No. of Shadow Pixels in OB}}{\text{No. of Segmented Pixels in OB}} > 0.9 \right) \tag{12}$$

2.2.4 Traffic Density Estimation

Once all the shadow blocks have been eliminated, the remaining vehicle occupied blocks are used to estimate traffic density. Finally the percentage vehicle occupancy (P) of the road segment and each lane is calculated which gives a fair idea about the level of traffic in each lane as well as the whole road segment. P for a frame can be defined as

$$P = \frac{\text{No. of VOB per frame}}{\text{No. of BOI per frame}} * 100 \tag{13}$$

Using the percentage occupancy level P, the traffic level can be classified into light, medium and heavy traffic density by fixing the percentage ranges for the different categories. The percentage ranges used in this paper for traffic density classification of a frame is given in Table 1. Since there are no set value of percentage occupancy to define light, medium or heavy traffic. We have performed detailed experiments to generate the optimum ranges for the classification process.

Traffic Density	**% Range**
Light	$P < 40\%$
Medium	$40\% \leq P \leq 65\%$
Heavy	$P > 65\%$

Table 1: Percentage Occupancy Range

3. Results

In this section, we present details about the datasets used, quantitative and qualitative evaluations of the pro-

Video	HighwayI [21]	HighwayII [21]	Highway[27]	TrafficDB [4]
Sample Frame				
Number of Frames	440	500	1700	13062
Image Size	320x240	320x240	320x240	320x240
Illumination Conditions	Sunny	Sunny	Sunny	Overcast,Clear,Rain
Background Shadows	Yes	No	Yes	No
Moving Shadows	Long	Small	Small	Small

Table 2: Dataset used for Evaluations

posed method and comparison with state-of-the-art methods. Our technique has been implemented on Matlab, Intel i3 processor CPU 2.40 GHz with 4 GB RAM.

3.1. Dataset

A summary of the datasets used for the evaluation of the proposed technique is given in Table 2. The TrafficDB dataset was used for the comprehensive evaluation of the proposed method. It consists of 254 videos - 5 seconds each which have been annotated as light, medium, and heavy traffic respectively. The other datasets were used for frame level evaluations.

3.2. Qualitative Results

In Fig. 4 the qualitative results of the proposed approach have been presented. Every image is annotated with the percentage occupancy of each lane and the total percentage occupancy. Finally it is classified into light, medium, heavy category using the total occupancy. The lanes have been numbered from left to right, i.e. lane 1 is the leftmost lane. Column 1,2 and 3 shows images from HighwayI, TrafficDB day and TrafficDB night video sequences respectively. In Column 1, the robustness of the proposed approach in differentiating vehicles from shadows can be visualized. Column 2 and 3 show detection results in varied illumination conditions. Fig. 4b, Fig. 4e,f and Fig. 4c,h,i show results from clear, rainy and overcast conditions respectively. It can be visualized that the proposed method is invariant to illumination conditions. HighwayI and Trafficdb videos have a huge difference in their camera angles. Thus, in addition to the robustness in dealing with shadows and illumination changes, the invariance of the method to camera perspectives is also evident.

It can be seen that the percentage occupancy reduces for all three categories for night time videos. This can be attributed to the fact that only part of the vehicle near the headlight gets detected and also, the safe distance between vehicles is considerably higher. Thus, the thresholds given in 1 have to be adjusted for night time detection.

3.3. Quantitative Results

For the quantitative evaluation, we have created the ground truth for the Highway, HighwayI and HighwayII videos. In each frame the vehicle occupied blocks have been annotated. In addition to that to evaluate the performance of shadow block elimination, the cast shadow blocks have also been annotated. For the quantitative evaluation of our proposed technique, the following parameters were calculated after comparing the detection results with the ground truth data.

- TP_S = No. of shadow blocks classified correctly.
- FN_S/FP_V = No. of shadow blocks classified as vehicles.
- FP_S = No. of vehicle blocks classified as shadows.
- TP_V = No. vehicle blocks classified correctly.
- FN_V = No. of vehicle blocks classified as shadow/background.
- TN_V = No. of shadow/background blocks classified correctly.

3.3.1 Shadow Block Elimination Evaluation

To test the robustness of shadow block elimination in the proposed approach, we have used the performance evaluation metrics from [21]. The authors proposed two metrics for moving shadow detection evaluation: Shadow Detection Rate η and the Shadow Discrimination Rate ξ,where subscript S is for shadows and V is for vehicle. Prati *et al.* have defined ξ using foreground, since for traffic surveillance, foreground is vehicles, vehicle has been used instead of foreground.

$$\eta = \frac{TP_S}{TP_S + FN_S}; \xi = \frac{TP_V + FN_V - FP_S}{TP_V + FN_V} \qquad (14)$$

For our approach, we have calculated these values at the block level. A high value of η = 96.56% and ξ = 98.68%

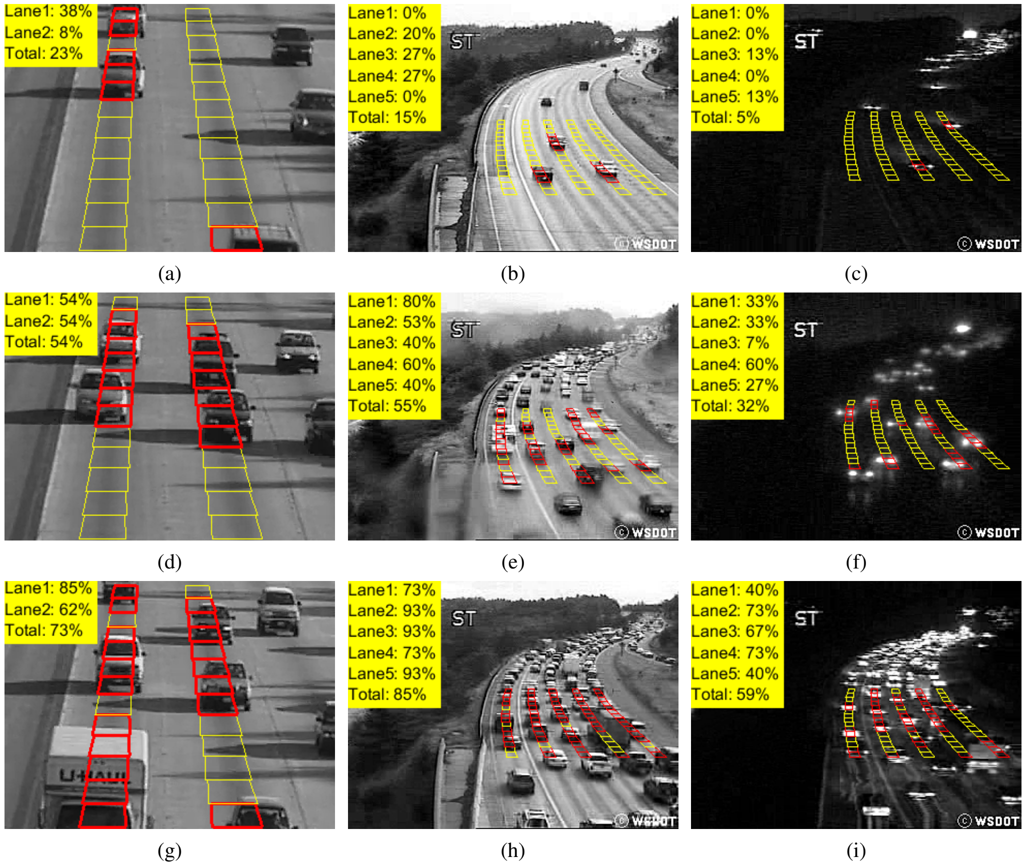


Figure 4: Percentage Occupancy Results for frames from HighwayI TrafficDB(Day) and TrafficDB(Night) in Column 1, 2 and 3 respectively. Row 1, 2 and 3 show Light, Medium and Heavy traffic respectively.

was achieved for HighwayI video shows the robustness of the shadow elimination technique used. Most of the mis-classifications occur when foreground objects have similar texture as the background [11][14].

3.3.2 Vehicle Block Detection Accuracy

In order to evaluate, the robustness of vehicle block detection, we calculate the True Positive Rate (TPR) and False Positive Rate (FPR) for the Highway, HighwayI and HighwayII. The TPR and FPR can be defined as

$$TPR = \frac{TP_V}{TP_V + FN_V}; FPR = \frac{FP_V}{FP_V + TN_V} \tag{15}$$

Table 3 presents the TPR and FPR values for the videos. The high value of recall i.e. TPR and low value of False Alarm Rate i.e. FPR represents the robustness of the presented approach in detecting vehicle blocks.

It should be noted the TPR and FPR for HighwayII are higher and lower than HighwayI respectively. One difference between the two videos is that the cast shadows in HighwayII are smaller than HighwayI and never reach the BOIs in the adjacent lanes. Thus, there are no false detection of shadow blocks as vehicle blocks in HighwayII leading to a low FPR. Another difference is the distinctive texture of the background for HighwayII which reduces the mis-classifications of vehicles as shadows.

Video	TPR	FPR
HighwayI	96.47%	0.42%
HighwayII	99.47%	0.11%
Highway	97.13%	0.65 %

Table 3: Vehicle Block Detection Accuracy

3.3.3 Traffic Density Estimation Accuracy

The overall traffic density estimation evaluation of the proposed approach has been done on the TrafficDB dataset. The results from the proposed system have been compared to the ground truth. In our technique each frame is classified as light, medium or heavy. Thus to classify a 5 second video, we choose the category to which the maximum number of frames from the video sequence have been classified. Table 4 provides a confusion matrix for the proposed system. It can be seen that most of the mis-classifications take place in heavy and medium category. There are two main reasons for these mis-classifications - (i) The presence of big vehicles (trucks, buses, etc.) leads to an increase in the percentage occupancy. (ii) Slow moving vehicles causing heavy traffic but occupying lesser number of blocks leads to a reduction in percentage occupancy.

		Predicted		
		Light	Medium	Heavy
Actual	Light	165	0	0
	Medium	3	37	5
	Heavy	1	7	36

Table 4: Confusion Matrix for TrafficDB

Table 5 presents the comparison of the proposed technique and other state-of-the-art techniques for video classification which were evaluated on the same database. Our proposed system achieves comparable accuracy to the existing methods, and achieves better accuracy than the method which uses only microscopic parameters [2].

Method	**% Accuracy**
Dynamic Texture Method[4]	94.50%
Spotiotemporal Gabor Filetrs[13]	91.50%
Spotiotemporal Orientation[9]	95.28%
Microscopic Parameters[2]	86.00%
Macroscopic Parameters[2]	95.28%
Symbolic Features[8]	96.83%
Motion Vector Statistical Features[22]	95.28%
Proposed Method	93.70%

Table 5: Traffic Density Estimation Accuracy

3.3.4 Run-Time Comparison

In this section, we compare the average time taken to classify a video from the TrafficDB dataset i.e. a 5 second video with 51 frames on average. Average time taken to process a video has been calculated for the proposed approach. Due to lack of publicly available implementations for the existing methods, the runtimes mentioned in the literature for different processors have been reported in Table 6. It can be seen that we are able to achieve comparable runtime to a GPU implementation of a Dynamic Texture Model.

The runtime reduction as compared to the existing methods can be attributed to the nature of computations in the proposed approach. Two main parts of the proposed method - background update and occupied block detection are mainly based on variance calculations for intensity blocks as opposed to pixel based analysis, this adds to a major reduction in computational complexity. The shadow elimination technique based on NCC[14], which is computationally complex, has been sporadically used on limited pixels. It has also been modified such that it can use look up tables for log calculations making it suitable for implementation on embedded platform. Also, being a block based method, our proposed technique can also be processed parallely which would lead to further reduction in run-time. Owing to the low complexity of our method, it is safe to say that it can be ported on a low-cost hardware platform which can be used for real-time road traffic density estimation.

4. Conclusion

In this paper, we presented a lane-wise traffic density estimation approach for traffic monitoring systems. The proposed method incorporates continuous background update and occupied block detection using block-based variance calculations, which significantly reduces the computational complexity compared to existing approaches that rely on pixel based analysis. Experiments on two different traffic videos demonstrated that the proposed method performs efficiently irrespective of illumination conditions, shadow conditions and camera perspectives, gives real-time performance and has comparable accuracy with existing state-of-the-art techniques. In particular, we show that the runtime of the proposed method, which is executed on a desktop computer, is only marginally higher than an existing GPU implementation. We plan to extend the proposed work to detect accidents and stopped vehicles in order to provide a more holistic understanding of the monitored area.

Method	**Runtime(s)**	**Processor**
Dynamic Texture Method[4]	193	2.16 GHz dual core,1 GB RAM
Macroscopic & Microscopic Parameters[2]	119	2.16 GHz dual core, 1 GB RAM
Mixture of Dynamic Texture Models [12]	8.19	NVIDIA Tesla C2070 GPU, 448 cores, 5376 MB Memory
Proposed Method	12.5	2.40 GHz Intel i3, 4 GB RAM

Table 6: Average video classification time for TrafficDB with average number of frames = 50.

References

[1] A. Amato, M. G. Mozerov, A. D. Bagdanov, and J. Gonzàlez. Accurate moving cast shadow suppression based on local color constancy detection. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 20(10):2954–66, oct 2011.

[2] O. Asmaa, K. Mokhtar, and O. Abdelaziz. Road traffic density estimation using microscopic and macroscopic parameters. *Image and Vision Computing*, 31(11):887–894, 2013.

[3] N. Buch, S. Velastin, and J. Orwell. A review of computer vision techniques for the analysis of urban traffic. *Intelligent Transportation Systems, IEEE Transactions on*, 12(3):920–939, 2011.

[4] A. B. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 771–776. IEEE, 2005.

[5] Y. L. Chen, B. F. Wu, H. Y. Huang, and C. J. Fan. A real-time vision system for nighttime vehicle detection and traffic surveillance. *IEEE Transactions on Industrial Electronics*, 58(5):2030–2044, 2011.

[6] Z. Chen, T. Ellis, and S. a. Velastin. Vehicle detection, tracking and classification in urban traffic. *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 951–956, 2012.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Statistic and knowledge-based moving object detection in traffic scenes. In *ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493)*, pages 27–32. IEEE, 2000.

[8] E. Dallalzadeh, D. S. Guru, and B. S. Harish. Symbolic Classification of Traffic Video Shots. In *Advances in Computational Science, Engineering and Information Technology*, pages 11–22. Springer, 2013.

[9] K. G. Derpanis and R. P. Wildes. Classification of traffic video based on a spatiotemporal orientation analysis. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 606–613. IEEE, 2011.

[10] L. Z. Fang, W. Y. Qiong, and Y. Z. Sheng. A method to segment moving vehicle cast shadow based on wavelet transform. *Pattern Recognition Letters*, 29(16):2182–2188, dec 2008.

[11] A. Gawde, K. Joshi, and S. Velipasalar. Lightweight and Robust Shadow Removal for Foreground Detection. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pages 264–269. IEEE, 2012.

[12] F. Gómez Fernández, M. E. Buemi, J. M. Rodríguez, and J. C. Jacobo-Berlles. Performance of dynamic texture segmentation using GPU. *Journal of Real-Time Image Processing*, aug 2013.

[13] W. N. Gonçalves, B. B. Machado, and O. M. Bruno. Spatiotemporal Gabor filters: a new method for dynamic texture recognition. *arXiv preprint arXiv:1201.3612*, 2012.

[14] J. C. S. Jacques Jr, C. R. Jung, and S. R. Musse. Background subtraction and shadow detection in grayscale video sequences. In *Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on*, pages 189–196. IEEE, 2005.

[15] K. Jiang, A. Li, Z. Cui, T. Wang, and Y. Su. Adaptive shadow detection using global texture and sampling deduction. *IET Computer Vision*, 2013.

[16] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis. A survey of video processing techniques for traffic applications. *Image and Vision Computing*, 21(4):359–381, 2003.

[17] P. Kumar, S. Ranganath, H. Weimin, and K. Sengupta. Framework for real-time behavior interpretation from traffic video. *Intelligent Transportation Systems, IEEE Transactions on*, 6(1):43–53, 2005.

[18] Z. Liu, K. Huang, and T. Tan. Cast Shadow Removal in a Hierarchical Manner Using MRF. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(1):56–66, jan 2012.

[19] L. Maddalena and A. Petrosino. The 3dSOBS+ algorithm for moving object detection. *Computer Vision and Image Understanding*, 2014.

[20] R. Mao and G. Mao. Road traffic density estimation in vehicular networks. In *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 4653–4658. IEEE, apr 2013.

[21] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows: algorithms and evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(7):918–923, 2003.

[22] A. Riaz and S. A. Khan. Traffic congestion classification using motion vector statistical features. In *Sixth International Conference on Machine Vision (ICMV 13)*, pages 90671A–90671A–7. International Society for Optics and Photonics, 2013.

[23] A. Sanin, C. Sanderson, and B. C. Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern recognition*, 45(4):1684–1695, 2012.

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

[25] K. SuganyaDevi and R. S. N Malmurugan. Efficient Foreground Extraction Based on Optical Flow and SMED for road traffic analysis. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 1(3):177–182, 2012.

[26] C. Tsai and Z. Yeh. Intelligent moving objects detection via adaptive frame differencing method. *Intelligent Information and Database Systems*, 2013.

[27] Y. Wang, P.-m. J. Fatih, P. Janusz, K. Yannick, and B. Prakash. CDnet 2014 : An Expanded Change Detection Benchmark Dataset. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 387–394, 2014.

[28] Y. Zheng and S. Peng. Model based vehicle localization for urban traffic surveillance using image gradient based matching. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pages 945–950. IEEE, 2012.