Traffic Density Estimation From Highly Noisy Image Sources

Vipin Jain
Department of Computer Science and Engineering
Polytechnic Institute of New York University
vjain02@students.poly.edu

Ashlesh Sharma, Aditya Dhananjay, Lakshminarayanan Subramanian
Courant Institute of Mathematical Sciences
New York University
{ashlesh,aditya,lakshmi}@cs.nyu.edu

Word Count: 5062 words + 9 Figures = 7312 words

Abstract

In this paper, we address the problem of how to accurately estimate the traffic density of road segments from highly noisy image sources. Conventional traffic density estimation techniques from camera feeds typically rely on high quality images. Surprisingly, a large number of live feeds from traffic cameras in developing regions are highly noisy due to poor camera quality, poor maintenance, limited field of view, limited network bandwidth (to upload high quality images),blur, multiple reflections and poor illumination effects. We propose a density estimation algorithm which uses a combination of conventional image processing techniques and semi-supervised learning using pre-labeled data to achieve high accuracy with minimal training. Our algorithm supports two different modes of operation for day-time and night-time and is accurate under both settings. We have tested our algorithm based on several hours of real-time traffic feeds from noisy sources in Nairobi, Kenya and Rio De Janeiro, Brazil.

1 Introduction

Traffic density estimation is an important requirement for real-time traffic management. One of the classic ways to monitor traffic density on roads is to use traffic cameras, gather an continuous image feed and process them using image processing techniques. There has been a large body of image processing techniques that have been proposed for accurately estimating the road traffic density from image feeds. A common requirement across most of these techniques is that the input source is reliable and provides high quality image feeds.

In this paper, we address the problem of how to accurately estimate the traffic density of roads from highly noisy image sources. This problem is particularly rampant in developing regions due to three real-world factors which affect the quality of the collected image data. First, in many developing regions, the CCTV cameras that are often used for gathering traffic data are not of high quality and are also poorly maintained. Second, the network connectivity in many of these countries is highly limited which also limits the ability of traffic signals to report high quality images in real-time to a central server due to bandwidth constraints. Third, with the miniaturization of low-cost lenses, many traffic cameras have a limited field of view which affects quality. In addition to the sources being noisy, the images could also suffer from poor illumination effects, multiple reflection sources and blur effects from light sources (headlights).

The problem of traffic density estimation from noisy image sources has received little attention. With high quality images, it becomes relatively easy to both accurately estimate the traffic density and even determine the contour of individual vehicles. With noisy sources, it can sometimes be relatively hard even for humans to accurately characterize the contour of individual vehicles in an image.

In this paper, we propose a semi-supervised algorithm for determining the traffic density of a road from a noisy image source. Our algorithm uses a combination of manual training and classification coupled with conventional image processing techniques to determine the approximate traffic density. Our algorithm consists of four basic steps:

1. Given a set of traffic images from a single image source, we mark a polygon-area in the image to characterize the region of interest that represents a specific

road segment; here, conventional image processing techniques for automatically identifying road segments do not work well for our collection of noisy image sources. In addition, we calibrate basic geometry parameters corresponding to the camera position and angle of view.

2. We automatically classify images into day and night time and have developed two different image analysis mechanisms based on pixel intensity distributions within the road segment of interest. For day time images, we measure the gray scale intensity distribution to characterize the "emptiness" of the road to obtain a measure of traffic density. For night time, we use an adaptive threshold mechanism to filter blur effects, poor illumination and multiple reflection sources and then use white-scale distribution to characterize traffic density.

3. To capture the notion of distance, we use a graded measure to map pixels to different areas within the road segment and obtain a weighted measure for the pixel intensity distributions.

4. We label a small collection of images for day and night time images corresponding to different levels of congestion and use this training data to learn the relationship between the pixel intensity distribution and the actual traffic density.

In this paper, we consider several hours of traffic data from noisy traffic camera sources from Nairobi, Kenya and Sao Paulo, Brazil. Based on a detailed evaluation from these sources, we show that our algorithm can accurately measure the road traffic density in both night time and day time images with minimal training from the user.

2 Related Work

Traffic and vehicle density estimations are important requirements for traffic management. Different solutions have been proposed primarily determined from the heterogeneity of the involved source data. Source data is procured from different sources like abounded camera feeds [6] , video feeds [4] , probe vehicles [5] and loop detector systems. [7] detects and tracks vehicles in a pole mounted high quality (HQ) camera's field of view. It extracts image features and successively detects vehicles based on the inclusion of vehicle features in consecutive passes. It identifies and analyzes both day and night time traffic. The illumination problem is solved by utilizing provided varying illuminated vehicle images. Individual vehicle identification is highly dependent on the quality of the available source feeds. [6] focuses on vehicle detection on the basis of traffic data collected from HQ cameras mounted on an DLR [1] aircraft. Such an approach is altogether infeasible due to the significant amount of involved investment. Also, no information on the night time approach has been provided where external illumination factor's highly skew the involved vehicle identification characteristics. Cloudy day measurements also suffer from the involved skewness in collected images. [1] uses a Kalman filtering process to characterize vehicle movement characteristics from a frame differencing technique. The results are highly applicable only

[1]German Airspace Center, http://www.dlr.de

towards day time traffic estimation. [3] presents a novel realtime technique to identify vehicle base front from camera feeds and leverage it towards measuring highway traffic counts and vehicular speeds. To reduce the image noisiness, background subtraction technique is utilized. [2] uses a simulation based approach towards vehicle platoon identification. The algorithm is responsible for per second vehicle density estimation and then utilizing the density information for platoon recognition. [4] also uses video feeds and proposes a neural network method for traffic density estimation and vehicle classification. Vehicle identification is done using a background subtraction technique followed by a neural network model which utilizes available vehicle properties. [8] determines traffic density utilizing an hidden markov model. By utilizing an unsupervised clustering scheme known as AutoClass, the paper proposes to accurately handle the varying illumination problem. The unsupervised machine learning is shown to achieve an high accuracy.

3 Problem Definition and Challenges

Traffic density estimation is an important characteristic required towards congestion detection, traffic management and traffic forecasting. Various approaches like camera/video feeds, car sensor data, loop detector data, and vehicle re identification mechanisms are utilized for density estimation. We build our work on available camera/video feeds which are abundant in today's highway deployments, considering the cheap cost and wide availability. Camera/video feeds suffer highly from bad, noisy image quality which complicates traffic density estimation, in addition to that, night time illumination and camera's limited field of view distorts any evaluation for traffic count.

Various image processing techniques have been leveraged to quantify the traffic density estimation and have been successful with varying degrees. Most of the image processing techniques follow a widely known approach of background subtraction followed by vehicle identification. Vehicle identification differs in day and night approaches, where daytime vehicles can be recognized from pre collected sample vehicle images and vehicle matching across the traffic data. Night time traffic estimation becomes cumbersome due to additional factors like poor lighting and surrounding illumination. Poor lightning basically involves the camera's limited quality in night time imagery. The illumination however is produced from various extraneous factors like vehicle's reflecting surfaces, billboard reflections, overhead signage and tunnel lightning etc. The illumination becomes denser as one observes a far off point in the field of view as compared to a near point. Most of the related work in this space avoids the night time traffic density estimation, moreover noisy images are never considered for density estimation because of the difficulty in property/object isolation. Also, most of the work described in section 2 estimates traffic density in sparse traffic conditions neglecting the aggrandized night effects due to poor image quality and intensified illumination. This paper considers the above mentioned effects to be much superior in traffic density estimation due to their contribution towards the noisiness in the image. Fig 1 shows images of two congested highways, where all previously applicable methods suffer in estimating correct traffic densities. A camera's orientation also significantly affects the camera's field of view, which focuses on the road area under consideration.

Figure 1: Bidirectional traffic camera image

Previous approaches have focused on identifying vehicle characteristics such as headlights, base fronts and also image morphology to infer rectangular patterns representing vehicles. Nonetheless, congested traffic images makes it much more difficult to identify individual vehicle characteristics because of the difficulty in any of the above mentioned property isolation.

4 Density Estimation Algorithm

In this section, we will explain our density estimation algorithm based on image feeds from a noisy source. Our density estimation algorithm consists of the following key steps:

Calibration and Pre-processing: In the calibration step, we manually mark an area in the image using a simple polygon to represent the road segment of interest. Next, we calibrate basic geometric parameters to obtain the angle of view, height and distance to road segment. The manual calibration is performed once for any road segment marking a single direction of the traffic under consideration.

Pixel Mapping and Graded Measure: In the pixel mapping step, we derive a mapping between the pixel positions on the road segment mapping and the physical road layout. This allows us to derive a graded measure of traffic density that accounts for varying traffic density with distance. The graded measure is further used to evaluate a graded pixel count for day/night time estimation.

Day time detection: The rationale used for day time detection is to use the gray area in the image to identify regions within the road segment with no cars. For day time detection, we use differences in the gray scale pixel distribution with the marked segment to obtain a graded measure of density in the pixel scale.

Night time detection: The rationale for night time detection is to detect headlights. However, for noisy images, accurate identification of individual headlights is a challenging proposition (even for humans). Here, we first have to adaptively threshold images to reduce blur effects and multiple reflection effects. Next, we use a white pixel distribution within the marked region to obtain a graded measure of traffic density.

Graded measure transformation: We consider a small sample of day-time images and night-time images of varying levels of traffic density and manually mark the density levels for each of the images. Next, we use a simple learning algorithm that can map

H - Height of the camera
C - Starting point of field of view
$d + X_{max}$ - Actual End point of field of view
E - Observed End point of field of view
G - Actual road point under inspection
F - Observed position of G in the image
x_i - Actual distance of road point from C
$h_\triangle$ - Observed height of the complete road length in the image
h_i - Observed height of Point G in the image

from the graded pixel scale to the actual traffic density measure.

Our overall algorithm is a semi-supervised learning algorithm since it leverages manual training at certain critical steps to estimate the traffic density. Next, we describe each of these algorithm steps.

4.1 Calibration and Pre-processing

In noisy images, accurate identification of road segments is a challenging task. This is specifically true if the image has two-way road traffic where the two lanes are very close to each other in the image. In addition, the image may contain several other feature which may not be of interest (trees, billboards etc.). A simple approach to performing better density estimation is to isolate only that polygon that contains the road, and perform the image analysis on only that portion of the image that falls within this polygon. Once we identify the road segment, we estimate 3 basic parameters of the road segment: H is the height of the camera; d is the distance to the real-world road segment and X_{max} is the length of the road segment. We approximately estimate these parameters from the image feeds.

4.2 Pixel mapping and Graded measure

An example image taken from a camera is shown in 4(a). The image is a colored image which is converted to grayscale for analysis. Next subsection describes how to reduce the extraneous noise in the image. Here, we explain the road's image with the help of Figure 2. The figure shows the cross section of the road area under consideration. For simplicity in evaluation, only a single lane is considered. The legend below shows the actual and observed points considered on a road segment.

C and *D* are the endpoints of road under consideration. Δ*ACD* represents the field of view of the camera. There is a distance $d = Htan\theta$ which comes before the near point in camera's field of view and any $length > d + X_{max}$ is beyond the coverage of the camera. The camera's field of view and image clarity varies according to the camera image quality and its zoom level.*CGD* is the actual road which is projected on the image. When marking the road segment in the image, let p_{Delta} represent the distance in pixels between the beginning of the road and the end of the road segment in the image. In other words, p_Δ represents the projection of the entire segment *CD* on

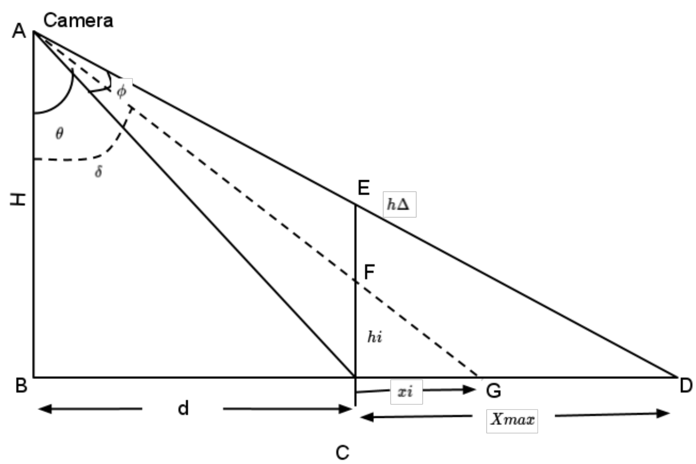


Figure 2: Field of View

the camera screen. While the camera scale is too small for geometric analysis, we can consider the projection *CE* as a virtual scaled-up projection of the image. Given that *CE* has a height of h_Δ, the scaling factor of the original image is h_Δ/p_Δ. Given any point *G* in the road, its corresponding projection point in the virtual scaled-up projection is represented by the point *F*. Let h_i be the height of *CF*. If p_i represents the distance of pixel corresponding to G from the beginning of the road segment, then we get: where

$$h_i = h_\Delta(p_i/p_\Delta)$$

This provides a mapping between the pixel position in the image and the actual distance in the real-world setting. As we go towards the far point of view, the density increases but the complementary effect on the camera's image is reduced. To solve this problem, we developed a density function based on the geometric properties of the road segment and its image. Overall Road traffic density is given by the product of a pixel count in row i represented by $count(i)$ and its corresponding weight function $W(i)$. We derive the density function as follows.

$$Density f(x) = \sum_i count(i) * W(i)$$

Based on geometric analysis, we can derive the appropriate graded weight to be: $W(i) = (x_i + d)/d$. Hence, we get the overall graded density function to be:

$$Density f(x) = \sum_i count(i) \frac{x_i + d}{d}$$

From the figure,

$$\frac{x_i}{h_i} = \frac{x_i + d}{H}$$

Solving the above two equations simultaneously provides us with:

$$Density f(x) = \sum_i count(i) \frac{H}{H - (P_i/P_\Delta)h_\Delta}$$

and

$$\frac{X_{max}}{X_{max} + d} = \frac{h_\Delta}{H}$$

Hence, given H, X_{max} and d and the traffic image, we can estimate the graded measure of the pixel count for the road segment.

4.3 Day time estimation

The day time estimation is performed by evaluating the gray levels in the image. The thesis behind evaluating gray levels is that the amount of daylight present in the image affects every vehicle in a similar fashion irrespective of the color of the vehicle. An empty road appears completely gray in color, while a non empty road show significant variation. Once we choose the road segment region from the calibration phase, we convert the image into an grayscale image and analyze the distribution of gray pixels within the marked polygon of interest.

Two different outcomes can be gathered from this part, first being the differences in the traffic density for congested and non congested road segments, the second being a non graded measure of the gray scale values present in the polygon of interest. The major difference between the congested and non congested segments can simply be visualized as the heavy contrast expected between a pure gray composition vs a lighter heterogeneous gray and white composition. This can be simply utilized as a naive measure of congestion detection in day time complementing the vehicle count estimation. A real world evaluation can simply utilize this to detect road congestion without performing a complete vehicle count analysis. Strictly speaking, the non graded measure can not be used in performing an exact vehicle count. But, we can leverage the techniques mentioned in the previous subsection for correctly identifying the number of vehicles present in the area of interest.

4.4 Night time estimation

While we use gray scale estimation for day time, we rely on white pixel scale estimation for night time. The white pixel is highly indicative of head-lights and the other image producing sources like billboards, overhead lights and signals etc. Depending on the white pixel count/density, a simple analysis can be performed for detecting the congestion in the image. As mentioned earlier, it can be computationally easier to only perform the final vehicle density count as against a full analysis over the pixel count and hence vehicle count. However, there are multiple factors which add an additional level of complexity in performing vehicle counting. Night time estimation is polluted by major factors including blur, reflections and illumination effects. To address these challenges, we adopt an adaptive thresholding process for segmenting the image and reducing noise. By varying the thresholding levels, one can remove most of the external noise in an image. The variation of thresholding levels helps to make a strict decision regarding the amount of external noise to be considered in the image. Figure 3 and Figure 4 shows the varying level of congestion in the two figures, and how well the varying thresholding levels are able to reduce external noise. The thresholding process is variable per traffic camera because each camera receives a different amount of light reflections. From the analysis it was found, that a thresholding value of 0.8 is successful in removing a significant portion of noisy illumination (as shown in Figure 3 and Figure 4). Although, a thresholding value of 0.8 is able to cover most of the noisiness, there have been multiple road instances where thresholding values of 0.7 and 0.9 were more apt.

After removing the overhead noise, we perform a graded count of white pixels across the road segment to have a overall score of headlights. As the pixel counter transitions from near point to the far point of field of view, we use the weight $W(i)$ as a varying density function to overcome the headlight size decrease with distance. The derivation of weight function $W(i)$ has been discussed in previous subsection.

4.5 Graded measure transformation

Converting from the graded gray pixel scale or the graded white pixel scale to the actual density value is not a simple linear transformation. In fact, the density is inversely correlated to the gray scale graded measure during day time and is directly correlated with graded white pixel scale during night time. We identify an supervised machine learning approach to evaluate the actual vehicle count from the graded white pixel count.

We explain the process for a single road traffic segment as the approach and parameters vary with the road segments under consideration. To perform the traffic density transformation, we considered a small sample of night time images of varying levels of density and performed an manual approximate vehicle count to measure the traffic density. After obtaining the two measures of vehicle count, from the graded white pixels and the actual vehicle density, we supervise the algorithm with the known image traffic density values. Specifically, consider a set of n total images where p images have been identified by their actual vehicle count. We give the graded white pixel count as g_i and the known vehicle density count as r_i for any image i. We define k and δ such that

$$k \in [g_i - \delta, g_i + \delta]; k \subseteq p$$

After obtaining k by choosing δ, we take an image i with a known g_i and an unknown r_i. A variable k is utilized for marking the k close neighbors of the image i in terms

(a) Original

(b) level = 0.6

(c) level = 0.7

(d) level = 0.8

Figure 3: Different Thresholding levels

of its graded pixel count g_i. With a dataset consisting of finite number of known and unknown points, we utilize the Shepard's method for evaluating the Inverse Distance weighting (IDW). The IDW takes all g_i into consideration as the finite known points and predicts the values of r_i. IDW is given by

$$r(u) = \Sigma_{i=0}^{k} \frac{w_i(u) * g_i}{\Sigma_{j=0}^{k} w_j(u)}$$

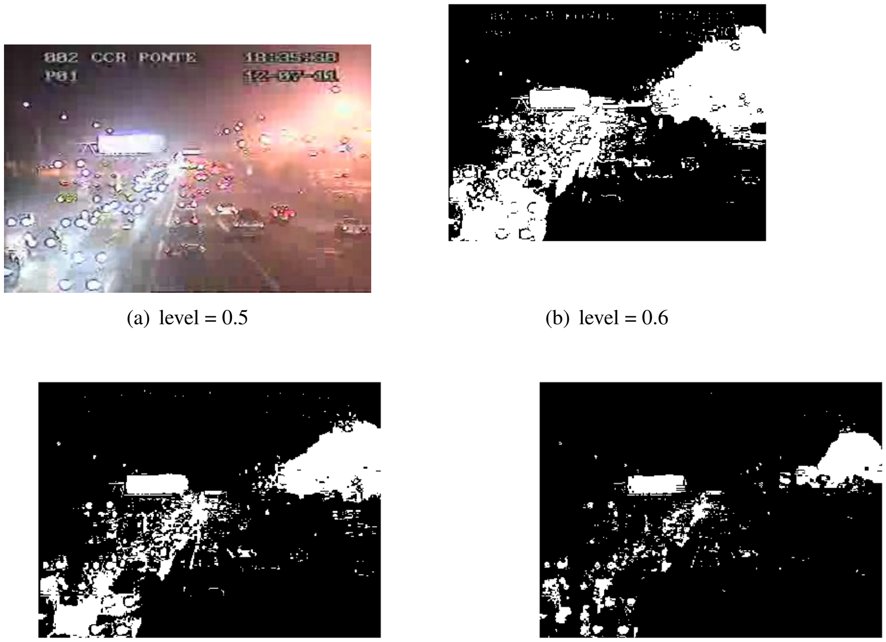
where u corresponds to an image with unknown actual density, and w_i corresponds to the weight function. and

$$w_i(u) = \frac{1}{\delta(u, u_i)}$$

P is an arbitrary positive integer known as the power parameter defaulted to two. Given the graded measure of a new image (either day or night time), we determined the training samples with the closest graded measure and computed the traffic density as a weighted average of these samples.

5 Evaluation

The source data was collected from two different sources, including Rio Niteroi Bridge in Rio de Janeiro (Brazil) and various traffic cameras across the Nairobi city (Kenya).



(a) level = 0.5

(b) level = 0.6

(c) level = 0.7

(d) level = 0.8

Figure 4: Different Thresholding levels

The image data from Rio Niteroi Bridge was preferred over the latter due to the former's camera height and field of view. Additionally, the amount of light captured in the latter images was of low intensity due to the high altitude. The former source provided video feed compositing from multiple video cameras mounted at various points on the bridge. The feed consisted of more than 7 cameras from which the images were extracted on a time variant basis. The cameras are mounted on top of highway overhead signs. With an average height of 5-7 m, the cameras record continuous sessions for the bidirectional road traffic. Figure 5 shows a polygonal region of interest, from the calibration and preparation stage. The area covered in the region of interest should be able to isolate vehicles as the far view of the camera blurs light to a significant aspect hampering vehicle identification

Figure 6 displays a sequence of images from the evaluation highlighting the semi supervised learning. We choose multiple images for the algorithm to identify a reasonable amount of graded pixel measure to actual vehicular density values.

Table 1 shows a small sample of the graded and actual pixel count identified to support supervised learning. As mentioned earlier, most of the parameters such as thresholding values, camera's field of view, vary for every single signal and road segment due to the high variation in extraneous light and camera's angle of orientation.

Figure 5: Polygonal Region of Interest

(a) g_i=35,r_i=44

(b) g_i=79,r_i=68

(c) g_i=70,r_i=60

(d) g_i=55,r_i=68

Figure 6: Night Time estimation: g_i=Number of Graded pixels, r_i=Number of Actual Vehicles

We utilize the measures derived from Section 4 and perform the evaluation for a single road segment. We keep the above mentioned constant evaluation parameters as $H = 5.5m$, $X_{max} = 200m$ and $d = 6m$, after the graded pixel measure and applying IDW, we evaluated the number of vehicles to be 52 in the right traffic lane. The real value estimated was 56 which shows the high accuracy of the approach. Figure 7 shows the traffic characteristics derived for the image.

As mentioned previously, the day time estimation was carried out on the basis of histogram analysis. During the day time, if a road is empty, a simple peak at a gray level is observed, showing absolutely no vehicle population. A similar methodology

White Pixels	Graded Pixel Count	Actual count
416	35	44
835	69	60
618	50	38
416	35	44
292	24	30
568	47	56
941	79	68
620	51	44
844	70	60
605	49	55

Table 1: White Pixel, Graded Pixel and Actual Vehicle Count

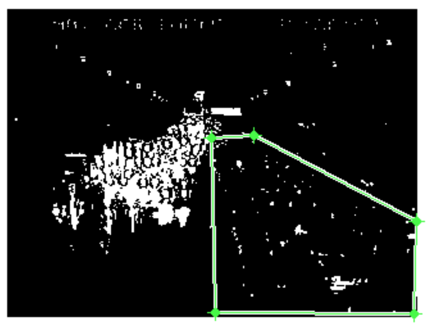


Figure 7: Night Time Estimation, Graded Pixel Measure($g_i = 47$), Actual Vehicle Count($r_i = 56$), Estimated Vehicle Count($r_i = 52$)

for day time estimation can be carried out from leveraging supervised learning and IDW.

6 Conclusions

Traffic density and vehicle density are utilized towards future traffic prediction. Not only these estimates help to solve realtime traffic congestion, they also help to forecast road/highway traffic characteristics. The paper focused on two strong factors which were understated in the previous research work in this area. Firstly, a method to analyze noisy images was outlined by reducing the noise utilizing simple thresholding process. Secondly, a night time traffic estimation method was proposed which compared to the day time estimation consists of extraneous noise contributors. A novel way of identifying traffic density was proposed in this paper giving weightage to the variation of the road section as projected in the image. We believe that our system

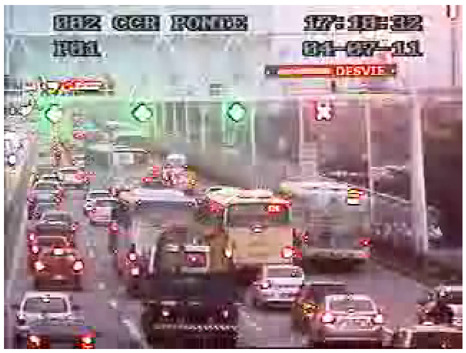


Figure 8: Day Time Traffic

Figure 9: Day Time Traffic Density Estimation Region of Interest

can dynamically monitor traffic density across any road segment based on the image source. Moreover, no additional cost is required considering the abundance of traffic cameras across highways. Such a system can be utilized for evaluating traffic density in real time to gauge changes in the traffic flow across a highway segment.

References

[1] P. Alcantarilla, M. Sotelo, and L. Bergasa. Automatic daytime road traffic control and monitoring system. In *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, pages 944 –949, oct. 2008.

[2] A. Gaur and P. Mirchandani. Method for real-time recognition of vehicle platoons. In *Transportation Research Record: Journal of the Transportation Research Board*, volume 1748, pages 8–17.

[3] N. K. Kanhere, S. T. Birchfield, W. A. Sarasua, and T. C. Whitney. Real-time detection and tracking of vehicle base fronts for measuring traffic counts and speeds on highways. In *Transportation Research Record: Journal of the Transportation Research Board*, volume 1993, pages 155–164.

[4] C. Ozkurt and F. Camci. Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. In *Mathematical and Computational Applications*, volume 14, pages 187–196.

[5] T. Z. Qiu, X.-Y. Lu, A. H. F. Chow, and S. E. Shladover. Estimation of freeway traffic density with loop detector and probe vehicle data. In *Transportation Research Record: Journal of the Transportation Research Board*, volume 2178, pages 21–29.

[6] P. Reinartz, M. Lachaise, E. Schmeer, T. Krauss, and H. Runge. Traffic monitoring with serial images from airborne cameras. In *ISPRS Journal of Photogrammetry and Remote Sensing*, volume 61.

[7] K. Robert. Video-based traffic monitoring at day and night vehicle features detection tracking. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, pages 1 –6, oct. 2009.

[8] E. Tan and J. Chen. Vehicular traffic density estimation via statistical methods with automated state learning. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 164 –169, sept. 2007.