A New Method for Traffic Density Estimation based on Topic Model

Conference Paper · December 2015

DOI: 10.1109/SPIS.2015.7422323

3 authors:

Razieh Kaviani Baghbaderani
University of Tennessee

Parvin Ahmadi
Sharif University of Technology

Iman Gholampour
Sharif University of Technology

A New Method for Traffic Density Estimation based on Topic Model

Razie Kaviani, Parvin Ahmadi
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
{r_kaviani, parvinahmadi}@ee.sharif.edu

Iman Gholampour
Electronics Research Institute
Sharif University of Technology
Tehran, Iran
imangh@sharif.edu

***Abstract*—Traffic density estimation is one the most important tasks for an intelligent transportation system (ITS). In this paper, a new framework for traffic density estimation based on topic model is proposed. This framework uses a set of visual features without any need to individual vehicle detection and tracking and automatically discovers the motion patterns in traffic scenes. Then, using likelihood value allocated to each video, traffic density could be estimated. Results on a standard dataset show high classification performance of our proposed approach and robustness to typical environmental and illumination conditions..**

***Keywords*—Traffic density estimation; Topic model; Intelligent transportation system.**

I. Introduction

Road traffic density estimation provides important information in Intelligent Transportation Systems (ITS) for road planning, intelligent road routing, road traffic control, network traffic scheduling, routing and dissemination [1]. Accurate calculation of traffic density is essential for the development of early warning and automatic signaling systems, statistics, planning and some security applications. Moreover, density data can be used to help drivers for choosing optimal way among variety of routes [2].

In the past decades, to estimate the traffic flow or traffic density on a road, magnetic loop detectors or supersonic wave detectors were used by road engineers. Present-day traffic management systems utilize image and video processing techniques to extract the same information from surveillance video of roads or junctions [3].

Video based traffic flow detection, as the most commonly used method in traffic flow detection, has a lot of advantages over the traditional methods. It can provide high-quality image information efficiently and stably. It is easy and economical to install video cameras. Besides, it would never damage the road, nor would it block the traffic. With the fast development of computer vision and digital image processing technology, video based traffic flow detection system has become increasingly robust, real time and intelligent. Due to these advantages, video based vehicle detection technology is becoming more and more important to ITS [4].

There are numerous researches completed on vehicle counting and traffic density estimation in the recent literature. Pang et al. [5] introduced a method for Vehicle Count in the Presence of Multiple-Vehicle Occlusions in Traffic Images. Chen et al. [6] developed a new system that counts vehicles in dark environments by using their headlights. Pornpanomchai et al. [7] studied on video vision for counting vehicles. Mohana et al. [8] studied on counting vehicles in real-time. Maduro et al. [9], succeed to find velocity of vehicles and traffic density in their study. Similarly, Wu and Gu [10] studied on finding velocity and density of vehicles but apart from Maduro et. al., their study was in real-time. Ozkurt and Camci [11] introduced another method based on neural networks that identify the vehicles and traffic density by processing the traffic videos. Zhao and Wang [12] studied on counting vehicles in hybrid traffic zones. Arora and Banga [13] used fuzzy based controller and morphological edge detection method. Their approach was based on the measurement of the traffic density by correlating the live traffic image with a reference image. The higher difference means the higher traffic density. Dangi et al. [14] proposed another method based on four lanes system in which, time is allocated according to the number of vehicles on the lane. The approach proposed by Gupta et al. [15] was based on computing the traffic load by comparing two images, the reference image and the live traffic image. Kanojia [16] suggested another method to control the traffic signal by using image processing. He first selects the reference image which is the image with no vehicles or less vehicles and every time matches real time images with that reference image. Based on the percentage of matching, traffic lights are controlled. But in this technique, image matching is performed by the edge detection.

In this paper, a new method for traffic density estimation is proposed. Density forecasting is done by using low-level features and applying topic models. The paper is organized as follows: Section 2 describes the background theory of topic models. In Section 3, we introduce our proposed method for traffic density estimation based on topic models. The experimental results are presented in section 4. Finally, section 5 concludes the paper.

II. Background Theory

Topic models have become quite popular due to their success in natural language processing. The most commonly used topic model is LDA. It is a generative probabilistic model and has become a popular model because it enforces a Dirichlet prior over the topic distributions and word distributions. In the following, we give a brief review of LDA:

In the video of N_d documents, each document is modeled as a mixture of K topics, where K is assumed known. Each topic k is modeled as a multinomial distribution over a vocabulary given by $\beta=\{\beta_k\}$. α is a Dirichlet prior on the documents. For each document d, a parameter θ_d of the multinomial distribution is drawn from Dirichlet distribution $Dir(\theta_d,\alpha)$. For each word w_{dn} in document d, a topic z_{dn} is drawn with probability θ_{dk}, and word w_{dn} is drawn from a multinomial distribution given by $\beta(z_{dn})$. α and β are the hyperparameters that must be optimized.

Given the parameters α and β, then the joint distribution of a topic mixture θ, a set of N topics z, and a set of N words w is expressed by:

$$P(\theta_d, z_d, w_d \mid \alpha, \beta) = P(\theta_d \mid \alpha)\prod_{n=1}^{N} P(z_{dn}, \theta_d)P(w_{dn} \mid z_{dn}, \beta). \qquad (1)$$

The marginal likelihood $P(w_n|\alpha, \beta)$ and hence the posterior distribution $P(\theta_d,z_d|\alpha,\beta)$ are intractable for expressing exact inference. Therefore, an inference method, such as variational Bayesian (VB) method must be utilized to approximate $P(\theta_d,z_d|\alpha,\beta)$ [19,20].

III. Proposed Method

Our traffic density estimation system includes four main components: ROI determination, feature extraction and construction histogram of words, model learning, and density estimation. The overall block diagram of the proposed system is illustrated in Fig. 1. The details of these components are explained in following subsections.

A. ROI Determination

The first step is to select region of interest (ROI) where the vehicle of interest road lane are present. The purpose of selecting ROI is to exclude the unnecessary background information such as other road lane. This unnecessary information is fixed in every frame of the live video because the camera is stationary.

B. Video Representation

The second step is the feature extraction in order to represent the video by features and construct the histograms of words. To perform the feature extraction, we first temporally divide the whole video into N_d non-overlapping short clips. Then, we utilize Shi and Tomasi [17] corner detector to find the key points and use these points to extract the optical flow using Lucas-Kanade method [18] from each pair of consecutive frames. A threshold TH_0 is applied to the magnitude of optical flow vectors to remove noise and only the reliable flows are preserved. In order to generate the visual words, the position and direction and magnitude of optical flow vectors are quantized. Optical flow vectors are denoted by (x, y, α, λ). The position (x, y) are quantized to the nearest position on a grid with spacing of H pixels and the angles of flow vectors, α, are quantized into N_m directions. Also, the magnitudes of flow vectors, λ, are quantized into N_c values. After spatial, directional and magnitude quantization, we obtain a vocabulary **V** of $N=N_a\times N_b\times N_m\times N_c$ visual words: $\mathbf{V}=\{v_i\}$, $i=1,\ldots N$, in which each word contains three aspects of contents, information about position, motion direction, and velocity of motion.

Visual words are accumulated over the frames of each video short clip to create the histogram of words. Then, a clip d_j of video $\mathbf{D}=\{d_i\}$ $i=1,\ldots N_d$, is represented as a vector $\mathbf{W}=\{w_n\}$ $n=1,\ldots N$, where w_n denotes the number of occurrence of word n in the clip. We use **D** as the input to the topic models in order to find correlations among these visual words.

C. Model Learning

The third step is model learning to learn the motion patterns in video clips. For this purpose, we use topic model, which was originally studied in the field of natural language processing. Topic model can capture word correlations in a set of textual documents to find latent topics. Therefore, we use topic model to discover the set of latent motion patterns from video by learning the distribution of visual features that co-occur, and to learn distributions of motion patterns that co-occur in the video. Then, we use these learned motion patterns to calculate likelihood measure to estimate traffic density in traffic videos.

D. Traffic Density Estimation

The next step is to calculate the traffic density in the desired target area. In order to determine the traffic density, we train the topic model with specific density such as light-density first and then estimate the traffic density by using log-likelihood measure at the end of the fitting phase. If the motion

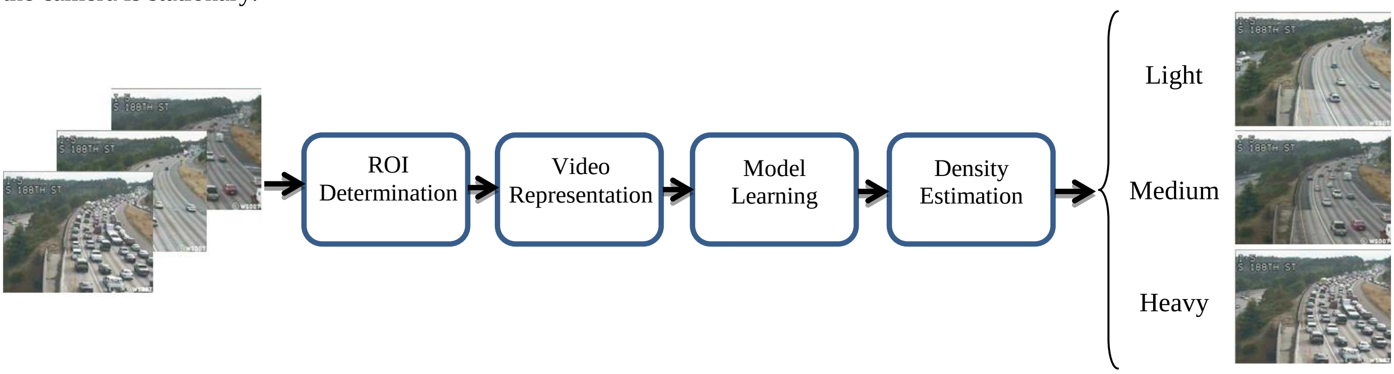

Fig. 1. Block diagram of a traffic density estimation based on Topic Model

patterns happening within the clip of test set correspond to those observed motion patterns in the training dataset, then the trained model should be able to find a suitable topic distribution explaining the bag-of-word representation of the clip. Thus, the clips with the same density as the training dataset will generally provide high log-likelihood. On the other hand, for the clips that include different density with training dataset, low likelihood will be achieved, because none of the learned topics is able to explain the observed visual words of that density. Since the likelihood is not normalized, this measure is highly dependent on the clip size. Therefore, to overcome this issue, normalized log-likelihood is used, that is derived by dividing log-likelihood of each clip by the number of visual words in that clip. Also, we use two thresholds to determine type of the traffic density such as light-density, medium-density and high-density.

IV. Experimental Results

We evaluated the performance of our proposed approach on UCSD database [21]. This is the database of highway traffic which was taken over two days from a stationary camera. This database contains 253 video sequences collected by Washington Department of Transportation (WSDOT) on interstate I-5 in Seattle, Washington. Each video clip has 42-52 frames and recorded at 10Hz with a frame size of 320×240 pixels. The videos in this database were clustered into three clusters: light traffic, medium traffic, and heavy traffic, in different weather conditions such as clear, overcast, and rainy. Example frames of this database are shown in Fig. 2. The details of the videos which results of them are presented in this paper are shown in TABLE I.

Since we are interested to estimate traffic density in only one road of highway, we define the region of interest and process on it. The target area is illustrated in Fig. 3 (b).

To create a visual word vocabulary, we followed the procedure in section 3(B). We divided each video clip into 1s length short clips. Also, each frame was spatially split into 10×10 cells and the motion direction was quantized into 4 orientations (up, down, left, right) and the velocity of motion was quantized into 4 values. Therefore, the sizes of the visual vocabulary were 32×24×4×4=12288 words. Figure 3(c) shows

Fig. 2. First row: different types of traffic density, a) light, b) medium, c) heavy; second row: different weather conditions, a) clear, b) overcast, c) rainy

TABLE I. The number of videos in used dataset

Total number of used videos	100
Number of videos for training	10
Number of videos for testing	90
Number of videos in clear weather	18
Number of videos in overcast weather	56
Number of videos in rainy weather	26

Fig. 3. a) traffic scene, b) ROI, c) extracted optical flow

this procedure which extract optical flows in ROI and quantize them.

Simulations were performed on laptop with Intel Core^{TM} i5-560 CPU 2.66GHz. In the experiments, we selected the optimal parameters for LDA and used 10 topics to model the scene activities. After extracting the topics in each video, we used log-likelihood and normalized log-likelihood to estimate the density. Fig. 4 shows the values of normalized log-likelihood for test videos.

To discover the two thresholds, the TH1 and TH2 were varied over the log-likelihood and normalized log-likelihood values to extract the ROC curves for light-density and heavy density. The ROC curves for these two densities are plotted in Fig. 5. From Fig. 5, it can be observed that normalized log-

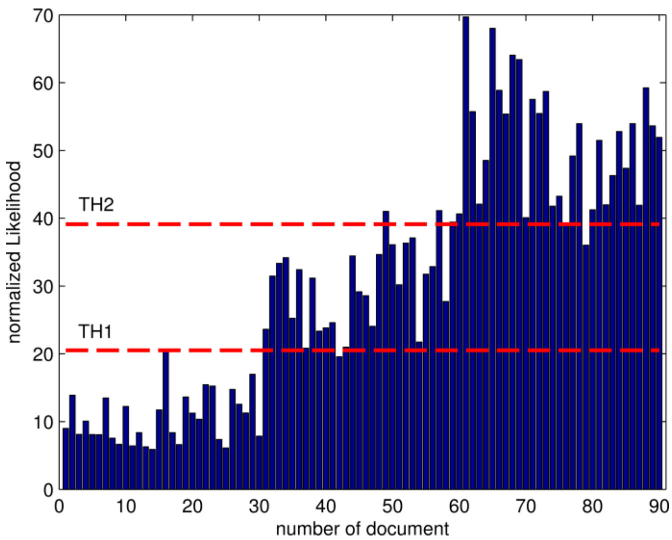

Fig. 4. Normalized log-likelihood values of test videos

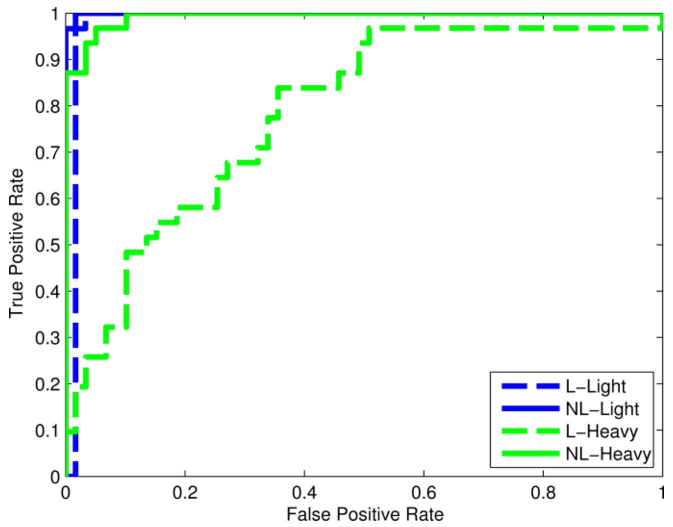

Fig. 5. ROC curves of light-density and heavy-density using log-likelihood (L) and normalized log-likelihood (NL)

likelihood (NL) achieved better performance than log-likelihood (L). According to the ROC curves for L, the best point for TH1 is where TPR=0.97 and FPR=0.017, and for TH2 is where TPR=0.84 and FPR=0.36. So TH1=20.52 and TH2=39.11 were selected. According to the ROC curves for NL, the best point for TH1 is where TPR=1 and FPR=0.016, and for TH2 is where TPR=0.96 and FPR=0.05. So TH1=20.52 and TH2=39.11 were selected to determine type of the traffic density.

To evaluate the capability of topic model for density estimation, every clip in the test dataset was labelled as light or medium or heavy according to their ground truth. Then, accuracy of the density estimation for each type was computed.

TABLE II shows the classification accuracy using used measures. Also, the time taken for feature extraction from each test video and for model learning and density estimation is reported in last column. As per Table II, the highest classification accuracy is obtained equal to 100% in light-density. The medium-density and heavy-density got 83.3% and 93.3% respectively. The proposed framework achieved an overall classification rate of 92.2%.

TABLE II. Classification results using LDA

	Light	Medium	Heavy	Time (ms)
Likelihood	97%	30%	87%	3320
N-Likelihood	100%	83.3%	93.3%	3424

V. Conclusion

In this paper, we have presented a framework to automatically classify complex traffic videos and determine their density, based on LDA. Experimental results have shown that this framework is able to accurately estimate the density of traffic videos even in bad illumination condition. In the future, we would like to confirm our results with more datasets and compare framework with other approaches.

References

[1] Ruixue Mao, and Guoqiang Mao. "Road traffic density estimation in vehicular networks." In *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 4653-4658. IEEE, 2013.

[2] Serap Kazan, Musa Balta, Kevser Ovaz, and Erman Ozer. "Traffic Intensity Estimation Using Real Time Computer Vision System.", 2013.

[3] Erhan, I. N. C. E. "Measuring traffic flow and classifying vehicle types: A surveillance video based approach." *Turk. J. Electr. Eng. Comput. Sci* 19 (2011): 607-620.

[4] Zhaoxiang Zhang, Yuqing Hou, Yunhong Wang, and Jie Qin. "A traffic flow detection system combining optical flow and shadow removal." In *Intelligent Visual Surveillance (IVS), 2011 Third Chinese Conference on*, pp. 45-48. IEEE, 2011.

[5] Chun Cheong Pang, William Wai Leung Lam, and Nelson Hon Ching Yung. "A Method for Vehicle Count in the Presence of Multiple-Vehicle Occlusions in Traffic Images." IEEE Transactions on Intelligent Transportation Systems, Vol. 8, No. 3, September 2007.

[6] Thou-Ho (Chao-Ho) Chen, Jun-Liang Chen, Chin-Hsing Chen and Chao-Ming Chang. "Vehicle Detection and Counting by Using Headlight Information in the Dark Environment." IIH-MSP '07 Proceedings of the Third International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007) - Volume 02, Pages 519-522. 2007.

[7] Chomtip Pornpanomchai, Thitinut Liamsanguan, Vissakorn Vannakosit. "Vehicle Detection and Counting From a Video Frame." Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008.

[8] H.S. Mohana M. Ashwathakumar G. Shivakumar. "Vehicle Detection and Counting by Using Real Time Traffic Flux through Differential Technique and Performance Evaluation." International Conference on Advanced Computer Control. Page(s): 791 – 795, 2009.

[9] Cristina Maduro, Katherine Batista, Paulo Peixoto and Jorge Batista. "Estimation Of Vehicle Velocity And Traffic Intensity Using Rectified Images." ICIP 2008. Page(s): 777 – 780, 2008.

[10] Jianping Wu, Caidong Gu. "The Design and Implementation of Real-time Automatic Vehicle Detection and Counting System." ICIECS 2009. Page(s): 1 – 4, 2009.

[11] Celil Ozkurt and Fatih Camci. "Automatic Traffic Density Estimation and Vehicle Classification for Traffic Surveillance Systems Using Neural Networks." Mathematical and Computational Applications, Vol. 14, No. 3, pp. 187-196, 2009.

[12] Rui Zhao, Xiaogang Wang. "Counting Vehicles from Semantic Regions." IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 2, June 2013.

[13] Madhavi Arora, V. K. Banga, "Real Time Traffic Light Control System", 2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012), pp. 172-176, Singapore, April 28-29, 2012.

[14] Vikramaditya Dangi, Amol Parab, Kshitij Pawar & S.S Rathod, "Image Processing Based Intelligent Traffic Controller", Undergraduate Academic Research Journal (UARJ), Vol.1, Issue 1, 2012.

[15] Pratishtha Gupta, G.N Purohit, Sweta Pandey, "Traffic Load Computation for Real Time Traffic Signal Control",

International Journal of Engineering and Advanced Technology, Vol. 2, Issue 4, April 2013.

[16] Sabya sanchi kanojia, “Real –time Traffic light control and Congestion avoidance system”, International Journal of Engineering Research and Applications (IJERA), pp.925-929, Vol. 2, Issue 2,Mar-Apr 2012.

[17] J. Shi, and C. Tomasi, “Good features to track,” Proc. of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 593-600, Seattle, Washington, June 1994.

[18] B. D. Lucas, and T. Kanade, “An iterative image registration technique with an application to stereo vision,” Proceedings of imaging understanding workshop, 1981.

[19] D. M. Blei, A. Y. Ng, M. I. Jordan and J. Lafferty, “Latent Dirichlet allocation,” Journal of Machine Learning Research 3 (2003) 993-1022.

[20] X. Wang, X. Ma, and W. Grismon, “Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(3):539-555, march 2009.

[21] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures" IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30(5), pp. 909-926, May 2008.

[22] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures" IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30(5), pp. 909-926, May 2008.