

Traffic Congestion Estimation using Video without Vehicle Tracking

Antony A. Siswoyo, Endra Joelianto, and Herman Y. Sutarto

Abstract—In this study, a traffic density monitoring system using the expectation maximization (EM) algorithm was tested on video data with varying traffic density levels. Experiments were conducted to find the most accurate image thresholding method to preprocess the images before they are fed to the EM algorithm. The algorithm successfully detected traffic density, separating it into two categories, namely ‘congested’ and ‘smooth’. Using the Bradley-Roth method for image thresholding produced the most accurate results.

Index Terms—Traffic control, Traffic sensor, Image thresholding, Expectation maximization, Congestion estimation, Vehicle tracking.

I. INTRODUCTION

THE monitoring of vehicle traffic flow is something people want to do in their daily lives to find out traffic conditions in advance [1]. Traveling activities like going to work or making a trip with family or friends need planning, for example related to the time required. If the public can know the traffic conditions in advance when they travel, their trip can be planned more optimally [2], [3]. However, traffic control systems in Indonesia, especially when monitoring levels of congestion, still depend on individual observation. These systems rely on information provided by traffic police officers through police communication media, internet-connected smartphone users, and so on. Hence, the traffic flow is determined as congested or smooth based on subjective judgment, as there are no definite guidelines for determining traffic density based on individual observation.

Driving habits of the majority of people in cities are often inconsiderate as exemplified by sudden lane shifts, two-wheeled vehicles passing through narrow spaces between two- or four-wheeled vehicles, vehicles cutting off other vehicles, public transportation making stops in non-designated places, and so on. These things make vehicle tracking programs ineffective when applied to city traffic. Therefore, several

researchers have conducted research into algorithms to estimate traffic conditions in cities using surveillance systems of traffic flow [4]-[6].

This paper presents a method to estimate traffic density without using a vehicle tracking method. The expectation-maximization (EM) algorithm [7]-[9] was used with four different image thresholding techniques [10]-[13] to find the most accurate one. Determining the image threshold can be seen as an extreme form of contrast enhancement, i.e., making bright pixels lighter and dark pixels darker, for the purpose of image-in-video extraction and recognition. The advantage of using the method under consideration is that it can be applied to images from almost any type of camera.

This paper is organized into five sections. The first section is the introduction, which briefly explains the background, objectives and contributions presented in this paper. The second section describes the different image thresholding methods that were tested in the experiments on video data representing varying traffic conditions. The third section explains the expectation-maximization method as well as the clustering method used to classify the average values of the binary images resulted from the image thresholding method. The fourth section discusses the experimental results by comparing the different image thresholding techniques, after which the results of the EM method are analyzed using the MATLAB program. The fifth section is the conclusion of this study.

II. IMAGE THRESHOLDING

Image thresholding is a method that aims to segment images by separating objects and the background based on differences in brightness or darkness. Before this method is applied to an image it must be converted to grayscale. When the image is raw, the intensity value of each pixel cannot be detected, because the value of each pixel is color coded. After the raw image has been converted to grayscale, the intensity values of all pixels can be compared to one another. The image threshold method uses these values to determine the image threshold value so that a binary image can be produced as the final result of this method.

A. Global Thresholding

The global thresholding method is a method of image segmentation based on the image’s histogram. As its name suggests, global thresholding applies a threshold to the image as a whole. One of the global thresholding methods that is often used as the basis for research is the global thresholding

A. A. Siswoyo is with the Instrumentation and Control Master Program, Faculty of Industrial Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia (email: anthonyssiswoyo@gmail.com).

E. Joelianto is with Instrumentation and Control Research Group, Faculty of Industrial Technology, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung 40132, Indonesia (corresponding author, email: ejoel@tf.itb.ac.id).

H. Y. Sutarto is with the Department of Intelligent System, PT. Pusat Riset Energi, Bandung 40226, Indonesia (email: hy.sutarto@rce.co.id).

technique introduced by Otsu [14]. The Otsu method is also used as learning material for image segmentation processing in programs such as MATLAB. Determining the optimal threshold is carried out in the following steps:

1. Separating objects and background by giving an initial threshold assumption value ($T = 1, 2, 3, \dots$).
2. Calculating the weight, average and variance values for the foreground and background for each threshold value used as an assumption.
3. Calculating the within-class variance value, which is the sum of the variance of the background and the variance of the object multiplied by their respective weights.
4. Then, the following three steps are iterated to get the minimal within-class variance value. When the within-class variance value has the minimal value, the threshold value that is used at that time is the optimal threshold value for the image.

B. Sauvola Method

In the Sauvola method [15], the calculation of the threshold value for the image is based on the local average value and the local standard deviation value around each pixel [16]. The local threshold value for pixel (x, y) can be calculated by:

$$T(x, y) = m(x, y) * \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right]_e \quad (1)$$

where $m(x, y)$ is the local average value and $\sigma(x, y)$ is the local standard deviation value. The parameter values in (1) are selected as $k = 0.5$ and $R = 128$ by way of 8-bit grayscale images, which are commonly used values when applying the Sauvola method.

C. Wolf Method

In the Wolf method [17], determining the threshold value is based on the average gray level of the image, the standard deviation around the pixel and the overall image, and the color level of the image. The threshold value is calculated as follows:

$$T(x, y) = (1 - k) * m + k * M + k * \frac{\sigma}{R} (m - M) \quad (2)$$

where k is 0.5, m is the local average value, σ is the standard deviation value, R is the minimum gray image level and M is the maximum gray level in the image.

D. Bradley-Roth Method

The Bradley-Roth method [18] uses the integral image value (sum-area table) to determine the threshold value. The steps taken to get the threshold value for each pixel are as follows:

1. RGB images or video data are converted to grayscale data or images.
2. The intensity of each pixel of the grayscale image is calculated and added up based on the location of each

pixel:

3.

$$I(x, y) = f(x, y) + I(x-1, y) + I(x, y-1) - I(x-1, y-1) \quad (3)$$

4. After the integral image value has been obtained, the sum of each function of the rectangle in the upper left corner (x_1, y_1) and the lower right corner (x_2, y_2) can be calculated using

$$\sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} f(x, y) = I(x_2, y_2) - I(x_2, y_1 - 1) - I(x_1 - 1, y_2) + I(x_1 - 1, y_1 - 1) \quad (4)$$

III. EXPECTATION-MAXIMIZATION

The last step of the proposed method is data clustering, which separates the traffic conditions observed into two categories, namely 'congested' and 'smooth'. The threshold value is determined using the 1D expectation-maximization method. This study used the 1D method because the average value obtained from a binary image is a 1D value.

Some of the steps that this method uses to perform its tasks are the following [7]-[9]:

1. The algorithm first gives the mean value and initial covariance, or initial assumption. This process starts with the E-step (expectation step).
2. After being given the initial assumption based on the mean value and the initial covariance, the probability of each of these points is calculated for one group using the following equation:

$$P(x_i | b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{(x_i - \mu_b)^2}{2\sigma_b^2}\right) \quad (5)$$

3. Then, the same value is calculated for each of the possible points of the blue group and the yellow group.

$$b_1 = P(b | x_i) = \frac{P(x_i | b)P(b)}{P(x_i | b)P(b) + P(x_i | a)P(a)} \quad (6)$$

$$a = P(a | x_i) = 1 - b \quad (7)$$

4. After the values have been obtained, the process enters the M-step (maximization step). In this stage, the mean and covariance values are recalculated using

$$\mu_b = \frac{b_1 x_1 + b_2 x_2 + \dots + b_n x_n}{b_1 + b_2 + \dots + b_n} \quad (8)$$

$$\sigma_b^2 = \frac{b_1 (x_1 - \mu_b)^2 + \dots + b_n (x_n - \mu_b)^2}{b_1 + b_2 + \dots + b_n} \quad (9)$$

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 + \dots + a_n x_n}{a_1 + a_2 + \dots + a_n} \quad (10)$$

$$\sigma_a^2 = \frac{a_1(x_1 - \mu_b)^2 + \dots + a_n(x_n - \mu_a)^2}{a_1 + a_2 + \dots + a_n} \quad (11)$$

5. The new mean and covariance values are used as input for the return E-step. This process is repeated until the value generated in the M-step no longer changes or has convergent values.

IV. EXPERIMENTAL RESULT AND ANALYSIS

A comparison between the four image thresholding methods discussed in Section II was conducted by looking at the results of the binary images and the averages of the binary images. After the most appropriate method for determining the image threshold was found, the process was continued by applying the EM algorithm to traffic video data with varying traffic conditions (dense, varied, and relatively quiet). The result of data clustering by the EM algorithm consists of data on the traffic flow density in the video.

A. Image Thresholding Experiments

These experiments were conducted to find the most accurate technique for determining the image threshold value to be combined with the EM method. The techniques used were the Otsu method, the Sauvola method, the Bradley-Roth method, and the Wolf method. The performance of each method was determined by looking at their ability to distinguish the background and the object.

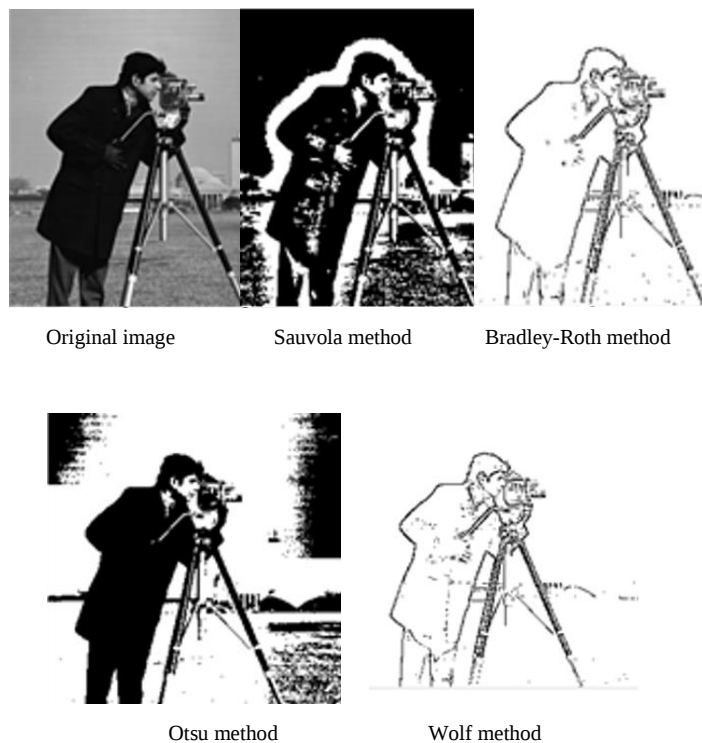


Fig. 1. Binary image result of each image thresholding methods.

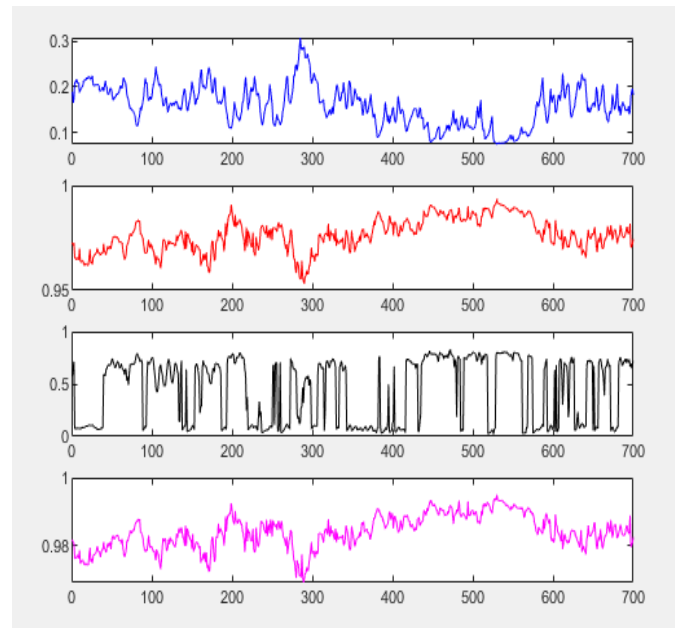


Fig. 2. Comparison of the mean values of each thresholding method. From the top: Bradley-Roth, Sauvola, Otsu, and Wolf.

Fig. 1 shows a comparison between the four image thresholding methods applied to the well-known cameraman image. Fig. 2 shows the mean values resulted from the respective methods. The capability of each method is indicated by the average value of the binary image. The Bradley-Roth method gave a more stable mean value than the others. For the Sauvola and Wolf methods the mean value was almost 1, which means that the binary image was almost completely white. The Otsu method had poor average value fluctuation. Therefore, the Bradley-Roth method was selected as the image thresholding method for further use in this study.

To compare the capability of forming binary images from each of the above image thresholding methods, this study also used a method to determine misclassification errors, which estimates the number of pixels that have classification errors (background or object). The ME value is between 1 and 0, where 0 means that the image is segmented accurately and 1 means that the result of image segmentation is very inappropriate. Thus, the smaller the ME value, the better the binary images produced by the image thresholding method. The ME values can be calculated by using the following equation:

$$ME = 1 - \frac{|B_G \cap B_S| + |F_G \cap F_S|}{|B_G \cap F_G|} \quad (12)$$

where B_G and F_G respectively represent the number of pixels of the background and the object in the actual image (ground truth image), and B_S and F_S respectively represent the number of pixels of the background and the objects in the results of the image thresholding method in the form of a binary image.

TABLE I
PERFORMANCE COMPARISON USING ME

Image	Sauvola Method	Bradley Method	Wolf Method
	0.61	0.004	0.61
	0.622	0.028	0.631

It was seen that the Bradley-Roth method was the most accurate method for processing video data and/or images of traffic flow. The ME values of the Bradley-Roth method were close to 0, producing a graph of average values of the binary image that matched the changes in traffic conditions.

B. Expectation-Maximization

Traffic density video data was captured in the area of Jalan Djuanda, Jakarta city, Indonesia. The video data had varying conditions, i.e., dense, varied, and relatively quiet. Analysis of these video data could show more specifically how each level of density is represented by the average value of the binary images resulted from image thresholding.

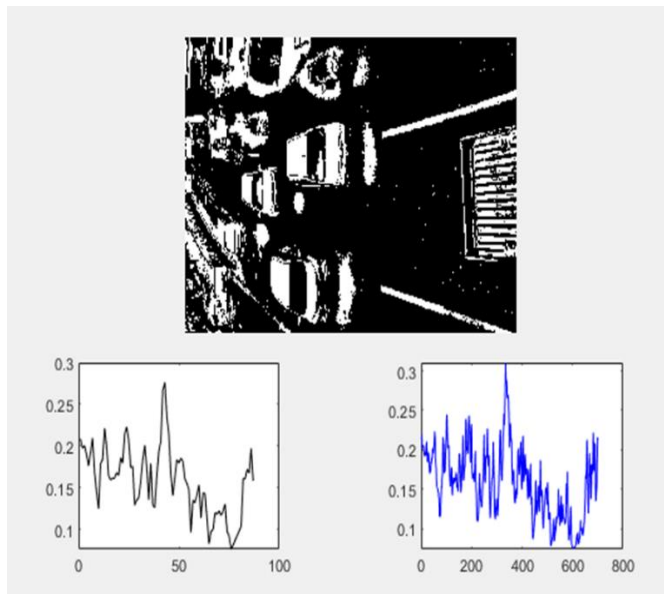


Fig. 3. Image thresholding results of (left) the mean value for every second and (right) the mean value for every frame in the traffic video.

From the results of the average values of the binary images, the three different types of traffic density mentioned above could be detected. Hence, it can be said that for video data with varying traffic conditions, the Bradley-Roth method proved its capabilities. When the average value of the binary images was in the range of 0.3 ± 0.02 , the road was in a congested condition. An average value of the binary images below 0.1 and even approaching 0 represents a condition where only one motorbike is close to the camera and two other vehicles appear in the distance.

After the average values of the binary images has been obtained, the values are forwarded to the EM algorithm. This process cannot be done in real time because the algorithm initially requires a lot of data. Therefore, it must be carried out in batches.

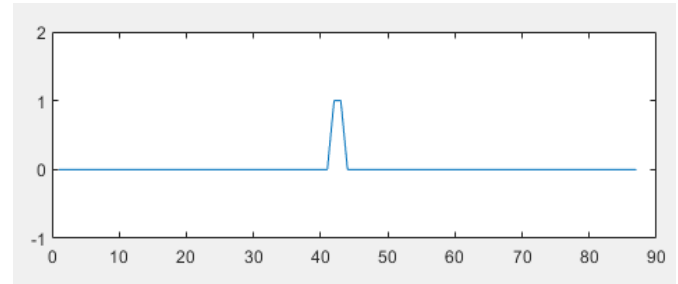


Fig. 4. EM clustering result.

Fig. 4 shows the result of data clustering by the EM algorithm into two groups, 'congested' and 'smooth'. The image shows that the EM algorithm successfully grouped the data that had converged values after iteration. Values from 0 to 0.25 mean that the value will be placed in the 'smooth' category, whereas when the value is higher, it is placed in the 'congested' category.

V. CONCLUSION

This paper considered a traffic density estimation system using any type of camera and the expectation-maximization method to group data into two categories, namely, 'smooth' and 'congested'. Experiments were done to find the image thresholding method that is most suitable for application to existing video data. The stability of the system in producing results in terms of 'smooth' or 'congested' traffic flow indicated that using the expectation-maximization method produced the expected results.

ACKNOWLEDGMENT

Endra Joelianto was supported by the ITB Research, Community Service and Innovation Program (P3MI-ITB) 2018, Ministry of Research, Technology and Higher Education of the Republic of Indonesia of Applied Higher Education Excellent Research University, Bandung Institute of Technology, Bandung, Indonesia. This study was partially supported by USAID through the Sustainable Higher Education Research Alliances (SHERA) Program – Centre for Collaborative Research (CCR), National Center for Sustainable Transportation Technology (NCSTT) under contract no. IIE0000078-ITB-1.

REFERENCES

- [1] K. Kuru, and W. Khan, "A framework for the synergistic integration of fully autonomous ground vehicles with smart city," *IEEE Access*, vol. 9, pp. 923-948, 2020.
- [2] C. Tang, J. Sun, Y. Sun, M. Peng, and N. Gan, "A general traffic flow prediction approach based on spatial-temporal graph attention," *IEEE Access*, vol. 8, pp. 153731-153741, 2020.
- [3] T. Jain, M. Johnson, and G. Rose, "Exploring the process of travel behaviour change and mobility trajectories associated with car share adoption," *Travel Behaviour and Society*, vol. 18, pp. 117-131, 2020.

- [4] F. Porikli and X. Li, "Traffic congestion estimation using HMM models without vehicle tracking," In *IEEE Intelligent Vehicles Symposium*, vol. 2004, pp. 188-193, June 2004.
- [5] M. T. Zulfikar, "Detection traffic congestion based on Twitter data using machine learning," *Procedia Computer Science*, vol. 157, pp. 118-124, 2019.
- [6] B. Pratama, J. Christanto, M. T. Hadyantama, and A. Muis, "Adaptive traffic lights through traffic density calculation on road pattern," In *2018 International Conference on Applied Science and Technology (ICAST)* (pp. 82-86). IEEE, 2018.
- [7] H. Y. Sutarto, R. K. Boel, and E. Joelianto, "Parameter estimation for stochastic hybrid model applied to urban traffic flow estimation," *IET Control Theory & Applications*, vol. 9, no. 11, pp. 1683-1691, 2015.
- [8] H. Y. Sutarto, E. Joelianto, and T. A. Nugroho, "Developing a stochastic model of queue length at a signalized intersection," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, pp. 2183-2188, 2017.
- [9] N. Sammaknejad, Y. Zhao, and B. Huang, "A review of the expectation maximization algorithm in data-driven process identification," *Journal of Process Control*, vol. 73, pp. 123-136, 2019.
- [10] T. Singh, S. Roy, O. Singh, T. Sinam, and K. Singh, "A New Local Adaptive Thresholding Technique in Binarization," *IJCSI International Journal of Computer Science Issues*, vol. 8, issue 6, no 2, November 2011.
- [11] M. Chandrakala, "Comparative Study and Image Analysis of Local Adaptive Thresholding Techniques," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 35, no. 9, pp. 423-429, 2016.
- [12] P. Wellner, "Adaptive Thresholding for the Digital Desk," Technical Report EPC-1993-110, 1993.
- [13] N. Senthilkumaran, and S. Vaithegi, S, "Image Segmentation by Using Thresholding Techniques for Medical Images," *Computer Science and Engineering: An International Journal*, vol. 6, no. 1, 2016.
- [14] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Transactions on System, Man and Cybernetic*, vol. 9, no. 1, pp. 62-66. 1979.
- [15] J. Sauvola, and M. Pietikainen, "Adaptive Document Image Binarization," *The Journal of the Pattern Recognition Society*, vol. 33, no. 2, pp. 225-236, 2000.
- [16] M. Sezgin, and B. Sankur, "Survey over Image Thresholding Techniques and Quantitative Performance Evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146-165, 2004.
- [17] C. Wolf, and J. M. Jolian, "Extraction and Recognition of Artificial Text in Multimedia Documents," *Pattern Analysis and Application*, vol. 6, no. 4, pp. 309-326, 2004.
- [18] D. Bradley, and G. Roth, "Adaptive Thresholding using Integral Image," *Journal of Graphics Tools*, vol. 12, no. 2, pp.13-21, 2007.