

## Estimation of Average Car Speed Using the Haar-Like Feature and Correlation Tracker Method

Muhammad Dzulfikar Fauzi\*<sup>1</sup>, Agfianto Eko Putra<sup>2</sup>, Wahyono<sup>3</sup>

<sup>1</sup>Master Program in Computer Science, FMIPA UGM, Yogyakarta, Indonesia

<sup>2,3</sup>Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: \*<sup>1</sup>[muhammad.dzulfikar.f@mail.ugm.ac.id](mailto:muhammad.dzulfikar.f@mail.ugm.ac.id), <sup>2</sup>[agfi@ugm.ac.id](mailto:agfi@ugm.ac.id), <sup>3</sup>[wahyo@ugm.ac.id](mailto:wahyo@ugm.ac.id)

### Abstrak

Kecepatan mobil yang melaju di jalan pada umumnya dapat diestimasi dengan menggunakan speed gun. Dibutuhkan upaya pemanfaatan CCTV (closed circuit television) sebagai alat yang dapat digunakan untuk mengestimasi kecepatan mobil sehingga dapat meringankan beban pihak penyelenggara jalan untuk mengestimasi kecepatan mobil. Penelitian ini membahas tentang estimasi kecepatan rata-rata mobil dengan metode Haar-like Feature yang digunakan untuk mendeteksi mobil, kemudian hasil deteksi di lacak menggunakan Correlatin Tracker untuk melacak perpindahan objek yang telah terdeteksi dan menghitung jarak perpindahan dari mobil, sehingga kecepatan mobil yang terdeteksi dalam video dapat diestimasi. Hasil dari estimasi rata-rata kecepatan dibandingkan dengan hasil pengambilan kecepatan dengan speed gun sehingga didapatkan error dengan pengujian MAE sebesar 5,55 km/jam dan simpangan baku yang dihasilkan adalah 4,61 km/jam, dengan demikian dapat disimpulkan sistem yang dibuat valid dan dapat digunakan oleh pihak penyelenggara jalan dalam memantau kecepatan rata-rata mobil.

**Kata kunci** — Haar-like Feature, Deteksi Mobil, Tracking, Estimasi Rata-rata kecepatan

### Abstract

The speed of a car traveling on the road can generally be estimated by using a speed gun. Efforts are needed to use CCTV (closed circuit television) as a tool that can be used to estimate the speed of the car so as to ease the burden on the road operator to estimate the speed of the car. This study discusses the estimated average speed of the car with the Haar-like Feature method used to detect the car, then the detection results are tracked using Correlatin Tracker to track the movement of objects that have been detected and calculate the distance of movement from the car, so that the speed of the car detected in video can be estimated. The results of the estimated average speed compared with the results of taking speed with a speed gun so that an error is obtained by MAE testing of 5,55 km / hour and the resulting standard deviation is 4,61 km / hour, thus it can be concluded that the system is made valid and can be used by road organizers to monitor the average speed of a car.

**Keywords**— Haar-like Feature, Car detection, Tracking, Average Estimation of Speed

## 1. INTRODUCTION

Discipline is important in driving, including regulating driving speed so as not to endanger other road users. One of the factors to determine the maximum vehicle speed limit can be seen in the speed limit indicator that is on the highway. Lack of supervision by officers who control driving speed limits results in motorists disregarding the driving speed limits. Speed measuring devices such as speed guns are useful for measuring the speed of vehicles passing on the road, however, the use of speed guns is still limited to watching vehicles that pass on the freeway. Based on the Law of the Republic of Indonesia Number 22 Year 2009 concerning Traffic and Road Transportation, the speed limit for the freeway ranges from 60 Km / hour to 100 Km / hour. As for the vehicle speed on urban roads not more than 50 Km / hour.

Vehicle detection is one of the important problems when trying to estimate vehicle speed, one of the methods that are often used is Haar-like Feature. The Haar-like Feature method is used to detect car plates [1] getting an accuracy of 94% with testing data of 3,000 car plate images. When it is used to detect the type of vehicle that is small, medium and large vehicles [2] get an accuracy of 77.8% when the road conditions are quiet and when the road conditions are busy accuracy becomes 28.8%. [3] Haar-like Feature is used to detect vehicles in the frame image on toll roads getting an average accuracy of 92.3%. Therefore it is necessary to have a tracking feature when the vehicle is successfully detected to reduce the error rate.

Research conducted by [4] and [5] use the same method namely Haar-like Cascade used to detect vehicles, the difference between the three studies is in the research of [4] Vehicle detection is used for Counting when the vehicle crosses imaginary lines in the frame, in research conducted by [5] the detection results are added with stereo sub-pixel matching to track vehicle movement which is then used to calculate vehicle speed. The results obtained when used for counting obtained an accuracy of 78% in the research of [4]. The results obtained when used to measure vehicle speed get an MAE of 2.46 in a study conducted by [5]. Machine learning in [6] and [7] research is used to estimate vehicle speed, each researcher uses different methods including a combination of SVM, CNN and deep learning. Research conducted by [6] classification and estimation of vehicle speed by utilizing a car front camera. Classifier Training is done by changing the image size to  $64 \times 64$  pixels and then extracting its features with the Histogram of Gradients to be used as a classification. The combination of them are Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM and Coarse Gaussian SVM. The best results from the combination that has been used is Cubic SVM with an accuracy of 94.29%. Whereas [7] use a deep neural network that is used for vehicle detection and tracking. Network architecture with 9 convolutional layers, 4 inception modules, one SPP layer and 2 fully connect layers. A deep neural network implementation can be implemented to detect and track vehicles in real time.

## 2. METHODS

### 2.1 Traffic

Defined as the movement of vehicles and people in the road traffic space, while what is meant by road traffic space is infrastructure intended for movement of vehicles, people and/or goods in the form of roads and supporting facilities. The government has a goal to realize traffic and road transportation that is safe, safe, fast, smooth, orderly and orderly, convenient and efficient through traffic management and traffic engineering. According to the Law of the Republic of Indonesia Number 22 the Year 2009 concerning Traffic and Road Transportation. Based on PERMENHUB No.111 of 2015. Vehicle speed limits are shown in Table 1.

Table 1 Speed limits and road types

No.	Road Type	Minimum	Maximum
1	Freeway	60Km/hour	100Km/hour
2	Intercity	-	80Km/hour
3	City	-	50Km/hour
4	Settlement	-	30Km/hour

## 2.2 Haar-Like Feature

Vehicle detection using Haar-like Feature, [2] Haar-like Feature is a classifier that is trained with several sample images of an object. The classifier is trained using the Adaboost algorithm. In this case, the image sample used is the image of the vehicle, in the form of the side view (left and right), front and rear. The size of the image used for training must be the same (for example 20x20), where later this will be a positive sample. Negative samples are images of different objects but still have the same size. This collection of images will produce a collection of object features called cascades. The classifier will produce a value of "1" if the inserted image contains recognized objects and the value of "0" if no objects are recognized. Figure 3 is some of the Haar features that are used to detect vehicles.

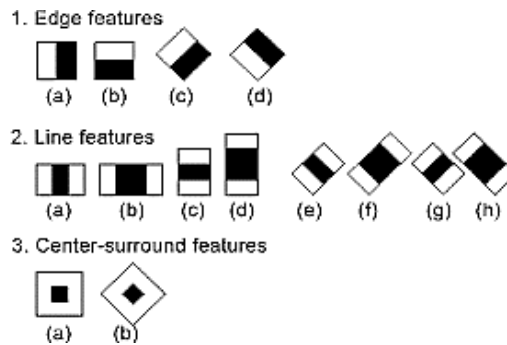


Figure 1 Haar Like Feature

The results of applying each feature to a particular image region are produced through the number of pixels located inside the black rectangle of the feature reduced by the number of pixels overlapping with the white rectangle. This rectangle is defined through the top left coordinates  $x$ ,  $y$ , width  $w$  and height  $h$ . The total pixels in the rectangular area of  $r_i$  are represented by  $RecSum(r_i)$ .

$$\begin{aligned}
 feature_{e_1} &= \sum_{i=1}^N W_i \times RecSum(r_i) \\
 &= \sum_{i=1}^N W_i \times RecSum(x, y, w, h)
 \end{aligned} \tag{1}$$

Equation (1) has  $N$ ,  $W_i$  and  $r_i$  values chosen randomly, depending on the object identified.  $RecSum$  is the sum of the intensities in each upright or rotated Haar feature inside the detection window.

In recognizing an object, cascade will ignore the image area that does not have an object that meets the criteria. This is very helpful in improving the performance of the classifier. In effect, the possibility of object detection errors decreases and detection accuracy increases.

## 2.3 Discriminative Correlation Filters for Multidimensional Features

The discriminative correlation filters method is commonly used for tracking, some filters are used in multi-dimensional correlation filters, which are using a 1-dimensional filter to

estimate scale, 2-dimensional filter for translation and 3-dimensional filter for localized space-scale localization of the target [8].

Map feature that is used to represent images on the d-dimension.  $f$  is a square patch on the target, extracted from the map feature. Dimension numbers are indicated by  $l \in \{1, \dots, d\}$  from  $f$  to  $f^l$ . The aim is to find the optimal filter correlation, consisting of one filter per dimension feature. can achieve optimal by minimizing the cost function, as shown in Equation (2)

$$\varepsilon = \left\| \sum_{l=1}^d h^l * f^l - g \right\|^2 + \lambda \sum_{l=1}^d \|h^l\|^2 \quad (2)$$

Variable  $g$  is the correlation result obtained from the training data with variable  $f$ . The parameter  $\lambda \geq 0$  is used to control the chaotic regularization term. Note to note is that Equation (2) only takes into account one training sample in case  $t = 1$ . As a solution for Equation (2) is shown in Equation (3).

$$H^l = \frac{\bar{G}F^l}{\sum_{k=1}^d \bar{F}^k F^k + \lambda} \quad (3)$$

The regularization parameter reduces the problem of frequently appearing zero in the spectrum, which will cause division by zero. Optimal filters can be obtained by minimizing output errors in all training patches. However, the solution requires a system of linear equations  $d \times d$  per pixel. To get a strong estimate it is necessary to update the numerator  $A_t^l$  and denominator  $B_t$  of the correlation filter  $H_t^l$  in Equation (4).

$$A_t^l = (1 - \eta)A_{t-1}^l + \eta \bar{G}_t F_t^l \quad (4a)$$

$$B_t = (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d \bar{F}_t^k F_t^k \quad (4b)$$

$\eta$  is the learning rate parameter. The correlation score of  $y$  in the  $z$  region of the feature map can be calculated using Equation (5). The targeted new state will be found by maximizing the score of  $y$ .

$$y = f^{-1} \left\{ \frac{\sum_{l=1}^d \bar{A}^l Z^l}{B + \lambda} \right\} \quad (5)$$

#### 2.4 Estimated Average Speed

the speed of the moving object can be calculated by finding the total distance traveled from the beginning of the moving object to the end of the moving object divided by the time taken to make the move [9], can be determined by using Equation (6) below

$$v = \frac{d}{t} \quad (6)$$

Where:

$v$  (velocity) = speed achieved in displacement.

$d$  (distance) = distance traveled.

$t$  (time) = the time taken to make the move.

Furthermore, based on the speed equation Equation (6), to get the results of the calculation of the speed of the vehicle from the results of the video recording, the distance value is obtained based on the distance of the reference taken from the beginning of the moving object to the end of the moving object that has been determined and for the time obtained from the number of frames needed when the object moves with the fps value of the video. Then it can be seen in Equation (7)

$$v = \frac{d \times d_p \times 3600}{1000} \text{ km/hour} \quad (7)$$

Where:

$d$  = The distance specified in the study.

$d_p$  = The number of frames for one second during video recording.

The process of finding moving objects in a sequence of frames is known as tracking. This tracking can be done by using object feature extraction and detecting moving objects or objects in the frame. By using the value of the position of objects in each frame we can calculate the position and speed of moving objects. The distance traveled by the object is determined using the center of the bounding box. Distances are calculated using the Euclidean distance formula. The Euclidean distance formula in two dimensions is the movement of a symbol between one point to another on the X-axis and Y-axis can be seen in Equation (8)

$$D_{(x,y)} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (8)$$

Where:

$D_{(x,y)}$  = distance traveled

$X_1$  = the position of point 1 on the X axis

$Y_1$  = the position of point 1 on the Y axis

$X_2$  = the position of point 2 on the X axis

$Y_2$  = the position of point 2 on the Y axis

### 2.5 Root Mean Square Error (RMSE)

MSE is used to measure the average square of errors or deviations and the calculation of root-based errors is RMSE also called the root mean squared deviation (RSMD). These individual differences are also called residues and RMSE functions to add them into a single measure of predictive power [10].

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{exp,i} - X_{est,i})^2}{N}} \quad (9)$$

Where:

N = Number of Samples

$X_{exp,i}$  = Actual value i

$X_{est,i}$  = Estimated value i

### 2.6 Mean Absolute Error (MAE)

MAE in its research is used to measure the normal size of the error in determining predictions, without considering the relationship. The normal MAE is used to test the total contrast between prediction and original perception where each individual difference has a weight. where  $X_{exp}$  represents the results of the experiment and  $X_{est}$  represents the estimated variable at the time i [10].

$$MSE = \frac{\sum_{i=1}^N |X_{exp,i} - X_{est,i}|}{N} \quad (10)$$

Where:

N = Number of Samples

$X_{exp,i}$  = Actual value i

$X_{est,i}$  = Estimated value i i

### 2.7 Standard Deviation

Standard deviation (SD) or also called standard deviation is a measurement of variability that best meets the requirements because it does not discard extreme values, calculate deviations from the average and account for account deviations from positive and negative values. Standard deviation can basically be said as the root of the sum of squares deviations divided by the amount of data in the distribution of values [9].

$$SD = \sqrt{\frac{\sum_0^i (x_i - \bar{x})^2}{N}} \quad (11)$$

Where:

N = Number of Samples

$x_i$  = Data i

$\bar{x}$  = Average

### 3. RESULTS AND DISCUSSION

This section discusses the results of the system that has been made along with testing the difference in error from the estimated vehicle speed obtained by the system compared with the vehicle speed data retrieval by using a speed gun tool.

#### 3.1 Effect of Video Size

Video size has a considerable influence on vehicle detection. The original size of the CCTV video obtained was  $1280 \times 720$ , in this research experiment the size changes can be seen in Table 2

Table 2 Effect of video size on processing speed

Size	Processing time per frame	FPS
$1280 \times 720$	0,494 seconds	2 fps
$854 \times 480$	0,383 seconds	3 fps
$320 \times 180$	0,087 seconds	11 fps



Figure 2 Detection results with frame size  $320 \times 180$

The video used in this study has a speed of 25 fps, which means one video has a processing speed of 0.04 seconds per frame. Thus the closest video is  $320 \times 180$ , as in Figure 2. However, with a small video size, the vehicle will be difficult to detect. While the size of  $1280 \times 720$  takes too long to do the process, the size used in this study is  $854 \times 480$ .

#### 3.2 Effect of ROI

ROI is a way of taking certain areas in an image that aims to make it easier to analyze and reduce the size of image storage. Video input used in this study is a two-lane road with four lanes, while only one lane left is taken to be analyzed. The process to get ROI is to close a portion of the video with white, can be seen in Figure 4.

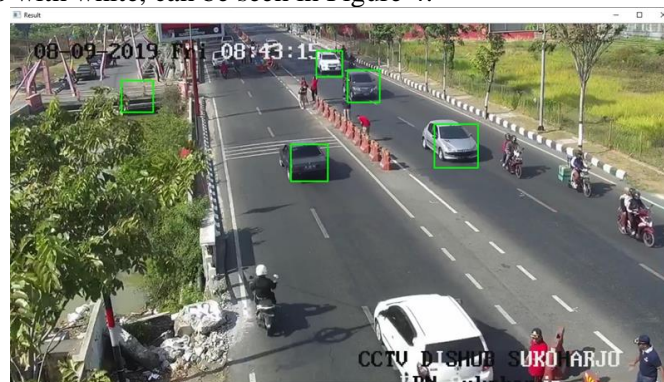


Figure 3 video before determining ROI

The vehicle detection results in Figure 3 show a false positive/false detection located on the far left detector. ROI determination is also used to take only the parts that are needed. As mentioned in the problem limitation, this study only uses one lane to estimate velocity, as in Figure 7.

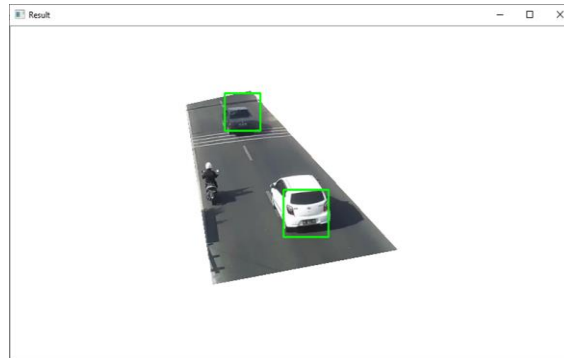


Figure 4 Video results after ROI is determined

The results of applying ROI to the video can be shown in Table 4. ROI used in this study aims to reduce detection errors, as in Figure 4 which at the beginning before using ROI there were errors in detection.

Table 3 The results of the influence of ROI

Video Input	Processing time per frame
Without ROI	0,280 seconds
ROI	0,131 seconds

ROI can also reduce the processing time which initially without using ROI takes 0.280 to 0.131 seconds for each frame so the performance has increased by 53.21%.

### 3.3 Tracking Testing

Tracking testing is carried out with two scenarios, namely without tracking and using tracking to estimate vehicle speed. testing in experiments without using tracking by giving two reference lines is used to detect when a vehicle is passing to record changes in time from the first and second reference lines. As in Figure 5.

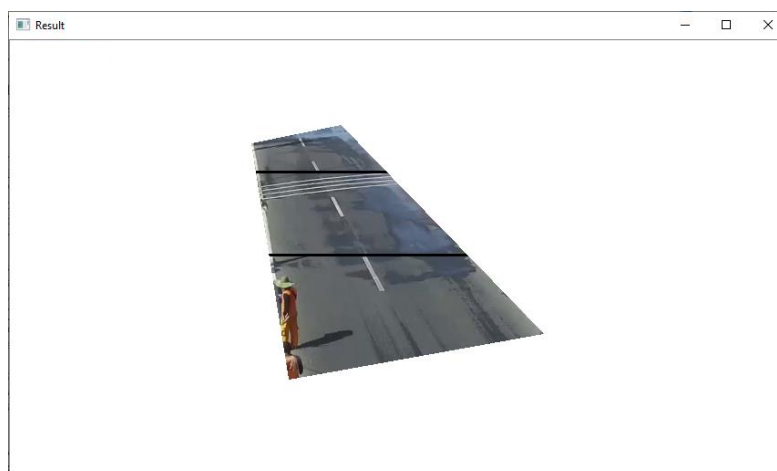


Figure 5 Reference Line



When a vehicle crosses the first (bottom) line, the crossing time will be recorded as the start time, when the vehicle crosses the second (top) line then the time will be recorded as the end time. From the results of the recorded time used as a reference separately to calculate the estimated speed of the vehicle. The reference line has an actual distance of 13 meters, for example, the time needed to cross is 1.25 seconds so to calculate the estimated speed to be as follows:

$$v = \frac{13}{1,25} = 10,4 \text{ m/s}$$

The results obtained are 10.4 meters / second, from these results are multiplied by 3.6 to make it to km/hour which is 37 km/hour. However, in experiments without using tracking there are some problems that occur such as the id of the vehicle being swapped can be seen in Figure 7 which causes an error that occurs very influential in calculating the estimated speed of a passing vehicle.

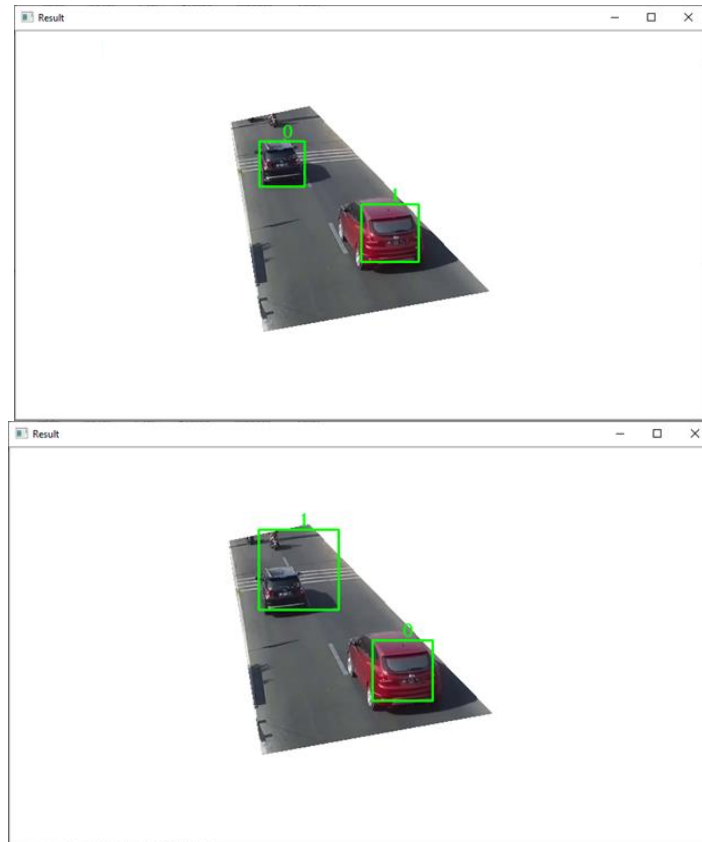


Figure 6 Exchangeable Vehicle ID

The swapping of the vehicle id can result in a wrong time recording the time the vehicle passes, as in Figure 8 the black car with the initial id 0 changes to 1 as well as that experienced by the red car with the initial id is 1 changes to 0. Thus a car with id 1 will cross the reference line faster than it should, for example, a car with id 1 that should cross two reference lines takes as much as 1 second can change to 0.15 seconds because the car id is exchanged with a car that is closer to the final reference line . Then the calculation of the estimated speed can be illustrated as follows.

$$\begin{aligned} v &= 13/1 \\ &= 13 \text{ m/s} \\ &= 13 \times 3,6 \\ &= 48,6 \text{ km/hour} \end{aligned}$$

The speed should be

$$\begin{aligned} v &= 13/0,15 \\ &= 86,7 \text{ m/s} \\ &= 86,7 \times 3,6 \\ &= 312,12 \text{ km/hour} \end{aligned}$$

The speed of the ID is swapped



Problems other than exchanged IDs that may occur in trials without tracking is when a detection fails when the vehicle is passing will result in an estimated speed of the vehicle being 0, because when the vehicle crosses the reference line one or two detection failures occur as in Figure 7.

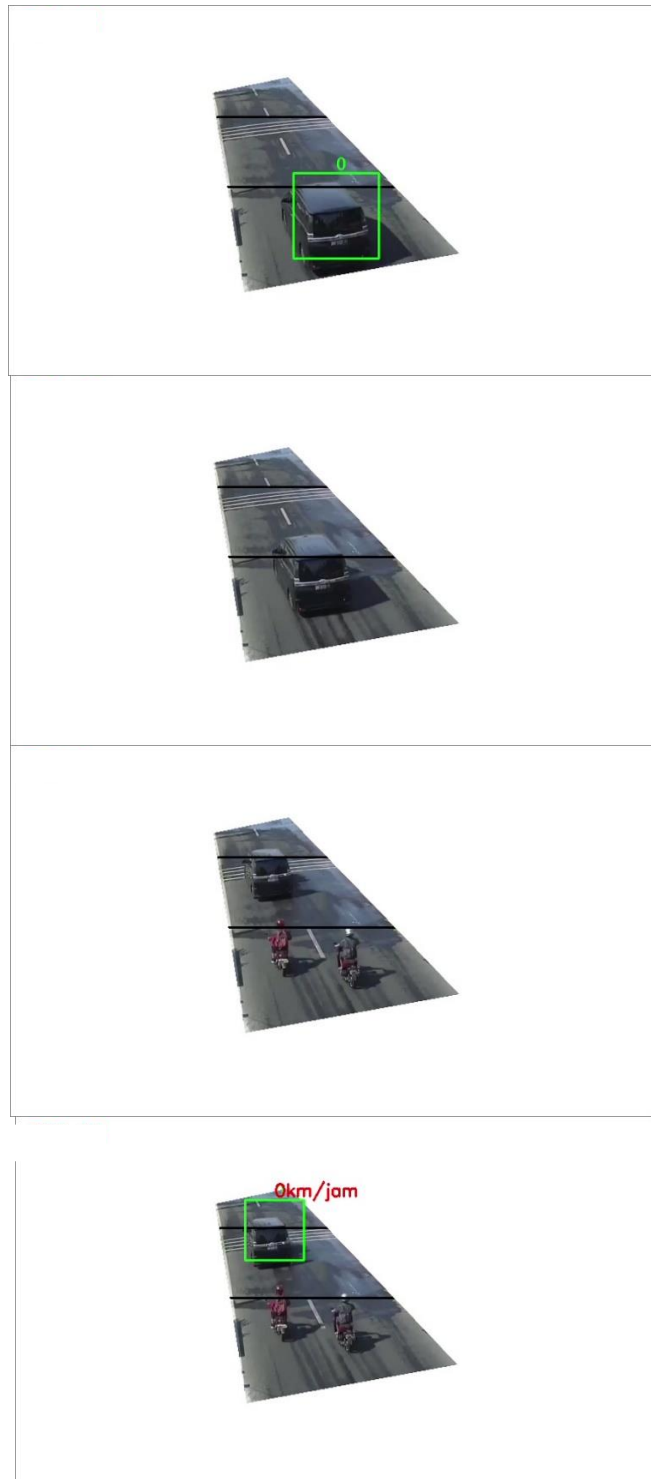


Figure 7 Vehicle failed detection

Table 4 Implementation of Speed Estimation Results

<b>With Tracking (A)</b>	<b>Without tracking (B)</b>	<b>Actual Speed (C)</b>	<b> A - C </b>	<b> B - C </b>	<b>(A - C)<sup>2</sup></b>	<b>(B - C)<sup>2</sup></b>
44	15	32	12	17	144	289
60	24	57	3	33	9	1.089
42	0	45	3	45	9	2.025
68	468	59	9	409	81	167.281
44	0	40	4	40	16	1.600
60	437	57	3	380	9	144.400
55	32	45	10	13	100	169
42	21	42	0	21	0	441
46	113	50	4	63	16	3.969
47	27	57	10	30	100	900
47	22	48	1	26	1	676
42	22	47	5	25	25	625
44	77	45	1	32	1	1.024
52	22	54	2	32	4	1.024
40	229	42	2	187	4	34.969
52	16	56	4	40	16	1.600
30	0	35	5	35	25	1.225
23	0	37	14	37	196	1.369
32	126	48	16	99	256	9.801
47	58	44	3	14	9	196

The median generated is based on Table 4 for |A-C| is 4 whereas for |B-C| is 34. The actual speed results are taken using a speed gun that is assisted by Sukoharjo DISHUB officer, the speed gun used in actual data collection has a fault tolerance taking the speed of approximately 2 km/hour, the speed taken is only vehicles that are driving at a constant speed or not braking.

### 3.4 Speed Evaluation

This stage is carried out to see how optimal the system has been made by testing to determine performance by using several methods, namely RMSE, MAE and Standard deviation. MAE and RMSE can be used together to identify variations in errors in a prediction set of average speeds on vehicles that have been recorded with the help of a speed gun. RMSE will always be greater or equal to MAE; the greater the difference between RMSE and MAE, the greater the variance in the error of each individual in the sample. If RMSE = MAE, then all errors on large scale have in common. Based on Table 4, the results of performance calculations can be seen in Table 5.

Table 5 Speed estimation test results

<b>Type Evaluation</b>	<b>With Tracking</b>	<b>Without Tracking</b>
RMSE	7,14 km/hour	138,87 km/ hour
MAE	5,55 km/ hour	78,90 km/ hour
Standard Deviation	4,61 km/ hour	11,75 km/ hour

The results of the test have quite a comparison between the use of the tracking and without tracking. According to [11] in his book for the standard deviation of error tolerance on urban roads with 2 lanes is 7.7 km/hour.

#### 4. CONCLUSIONS

The error tolerance limit for measuring vehicle speed is 7.7 km / hour according to [11] and the standard deviation generated using tracking when estimating the average speed is 4.61 km / hour so that it still meets the error tolerance limit.

The use of tracking in the estimation of the average vehicle has a good performance and can help detection when there is a failure of detection and car id that may be exchanged when more than one car crosses the video, so it can be concluded that the use of tracking can help the failure of detection on a car that is being drove off.

Research conducted by [Hua et all] with the theme of estimating vehicle speeds using deep learning resulted in an RMSE of 12,109 compared with research that had been conducted to obtain an RMSE of 7.14.

#### ACKNOWLEDGEMENTS

Thanks to DISHUB Sukoharjo who has been giving CCTV video data for this research.

#### REFERENCES

- [1] K. Zheng, Y. Zhao, J. Gu, and Q. Hu, "License Plate Detection Using Haar-like Features and Histogram of Oriented Gradients," *2012 IEEE Int. Symp. Ind. Electron.*, no. November, pp. 1502–1505, 2012.
- [2] A. Lazaro, J. L. Buliali, and B. Amaliah, "Deteksi Jenis Kendaraan di Jalan Menggunakan OpenCV," *J. Tek. ITS Vol.6*, vol. 6, no. 2, 2017.
- [3] M. Irfan, B. A. A. Subondo, and I. Candradewi, "Sistem Klasifikasi Kendaraan Berbasis Pengolahan Citra Digital dengan Metode Multilayer Perceptron," *IJEIS*, vol. 7, no. 2, pp. 139–148, 2017.
- [4] N. Mer and M. P. Bhamare, "Image processing based Tracking and Counting Vehicles," *2019 3rd Int. Conf. Electron. Commun. Aerosp. Technol.*, pp. 335–339, 2019.
- [5] M. Jalalat, M. Nejati, and A. Majidi, "Vehicle detection and speed estimation using cascade classifier and sub-pixel stereo matching," *Proc. - 2016 2nd Int. Conf. Signal Process. Intell. Syst. ICSPIS 2016*, pp. 1–5, 2017.
- [6] J. J. P. Belen, J. C. V. Caysido, A. B. Llena, E. J. O. Samonte, G. N. Vicente, and E. A. Roxas, "Vision based classification and speed estimation of vehicles using forward camera," *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, no. March, pp. 227–232, 2018.
- [7] X. Gu, Z. Chen, T. Ma, F. A. N. Li, and L. Yan, "REAL-TIME VEHICLE DETECTION AND TRACKING USING DEEP NEURAL NETWORKS," *2016 13th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process.*, pp. 167–170, 2016.
- [8] M. Danelljan, G. Häger, and M. Felsberg, "Accurate Scale Estimation for Robust Visual Tracking," *Proc. Br. Mach. Vis. Conf. BMVC*, p. 11, 2014.
- [9] N. H. Tsani, I. B. Dirgantoro, and A. L. Prasasti, "Impelementasi Deteksi Kecepatan Kendaraan Menggunakan Kamera Webcam dengan Metode Frame Difference," *e-Preceeding Eng.*, vol. 4, no. 2, pp. 2373–2381, 2017.
- [10] J. Sangeetha and V. S. J. Prakash, *Improved Feature-Specific Collaborative Filtering Model for the Aspect-Opinion Based Product Recommendation*. Springer Singapore, 2019.

- [11] A. Pande and B. Wolshon, *TRAFFIC ENGINEERING HANDBOOK*, Seventh Ed. Hoboken, New Jersey: John Wiley & Sons, Inc, 2016.