Vehicle Speed Calculation from Drone Video Based on Deep Learning

Tashfiq Rahman
Dept. of Computer Science
King Mongkuts University of Technology
Thonburi
Bangkok, Thailand
tashfiq.rahman@mail.kmutt.ac.th

Muhammad Arie Ladhika Siregar
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
arie16@mhs.te.its.ac.id

Arief Kurniawan
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
arifku@ee.its.ac.id

Susi Juniastuti
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
susi@te.its.ac.id

Eko Mulyanto Yuniarno
Dept. of Computer Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ekomulyanto@ee.its.ac.id

***Abstract*—The number of the vehicles continues to increase from year to year and cause congestion especially when there is a long holiday. The cost of installing and maintaining the recording equipment and supporting infrastructure is quite extortionate. Therefore, the calculation of speed with a flexible and inexpensive cost requires drones. The purpose of this study is to monitor the flow of traffic in areas where there is no supporting infrastructure to install traffic cameras used to calculate the speed of the vehicles during long holidays. The camera shots taken from the drones are used to detect the types of passenger cars using library of You Only Look Once (YOLO) deep learning. The detected object is then tracked by using SORT (Simple Online and Realtime Tracking) to always detect the same object and calculate the speed of the vehicle when passing the RoI (Region of Interest). On the 5, 10, and 15 meters RoI, the lowest error values are 4.79%, 4.38%, and 2.96% respectively. Each of the results was obtained at a height of 10 meters. The proposed method can prove the error calculation of speed vehicle is small when the frame rate (fps) is high, altitude is low and the distance ROI is long.**

***Keywords*—Vehicle Speed, Deep Learning, YOLO, Drone**

I. Introduction

Based on data from the Central Statistics Agency (BPS) Indonesia, the number of cars and other vehicles in Indonesia was continued to grow from year to year [1]. The increasing number of vehicles causes traffic congestion on the road to be high, especially during long holidays. Thus, traffic monitoring and control are needed to avoid traffic jams. One method of monitoring traffic is using a camera as a sensor. Previous studies used fixed cameras to estimate vehicle speed [2] [4]. Besides fixed cameras, speed vehicle estimation also used video cameras from drones [3]. Drone is used to take video data from a predetermined height by taking video vertically so that the top of the car will be seen in the video image. In the study [2] [3], the authors proposed the Background Subtraction method to detect vehicles in a region of interest (ROI).

Deep learning is one of the emerging research fields in technology in recent years. YOLO is a deep learning library that uses a convolution neural network (CNN) architecture. YOLO is implemented in real time applications. Vehicle speed measurement with the emergence of deep learning for object classification and detection is used in various applications such as supervision, traffic management, and rescue tasks of HPV (High Priority Vehicle) in determining the fastest route based on the use of vehicle speed data.

In this study, we proposed to calculate the speed of vehicles passing through the video area of drone. YOLO version 3 deep learning will detect objects that enter the camera and pass the Region of Interest (RoI). The object will be tracked and then measured its speed based on the frame when entering RoI and exiting RoI. From this paper, the system is based on deep learning to measure the speed of objects from drone video data. The rest of paper is organized as follows: Section 2 is the literature review which is the reference of the study; Section 3 describes our proposed method to estimate vehicle speed; some experimental results are shown in Section 4; and, finally, some conclusions are discussed in Section 5.

II. Literature Study

A. Speed Calculation

The calculation process to get the speed of the detected object that moves and passes the RoI [4] according to the equation 1. Speed is represented by v (Km/h), d_{RoI} is the distance of Region of Interest (m), fr is the framerate (fps), and n is the numbers of frame.

$$v = \frac{d_{RoI} \times fr \times 3600}{n \times 1000} \tag{1}$$

B. Drone

Drone is mini planes that operate without a pilot and fly independently guided by the Global Positioning System (GPS) on-board, stabilized with a gyro sensor and a 3 axis magnetometer mounted on the autopilot microchip and monitoring drone telemetry at the Ground Control Station (GSC). the aircraft is controlled at the Ground Control Station (GSC) using radio control and changes to be autonomous at a fixed height. The antenna at the GSC will send a communication link to the drone and the drone will send back to the GSC by sending location, height, and speed [5].

C. CNN

The concept of Deep Learning is one type of artificial neural network algorithm that is processed using several

978-1-7281-8283-4/20/$31.00 ©2020 IEEE

hidden layers of non-linear transformation of input data to calculate the output value. There are several types of deep learning, one of which is a Convolution Neural Network (CNN). CNN is based on the visual cortex in humans and is the neural network of choice for computer vision and video recognition. CNN consists of a series of convolution and sub-sampling layers followed by a fully connected layer and layer normalization. Some of the convolution layers extract feature that is increasingly revised at each layer that moves from the input to the output layer. The fully connected layer which carries out the classification follows the convolution layer. Subsampling or pooling layers are often inserted between each convolution layer. Each layer consists of 2D neuron groups called filters or kernels [6].

D. You Only Look Once (YOLO)

YOLO is a method of object detection in images that are based on Convolutional Neural Network (CNN) [7]. YOLO already has several versions in development, namely: YOLOv1, YOLOv2, and YOLOv3. YOLOv3 divides the input image into 13×13 grids, each grid is responsible for detecting an object on the grid. The Neural Network YOLOv3 uses regression to predict the object value per box [8] that predicts 4 coordinates for each bounding box [9]. Each bounding box will overlap with other bounding boxes, if the next bounding box exceeds the threshold, the prediction will be ignored. Objects are offset from the top left corner of the image and the previous bounding box. To predict each class, bounding boxes use multi label classification. The first step is data normalization. Then, separating the training data and test data into 80 percent as training/calibration data and 20 percent as test/validation data. This composition is adjusted to the standard input from Ignore. The next step is data classification. In this step, the new model is built using only calibration data and tested with validation data. The output of this process is the calibration prediction, the level of confidence of each label in rational numbers between 0 to 1, the validation prediction, and the evaluation results [11].

E. SORT

Simple Online and Realtime Tracking (SORT) is an algorithm that uses the Kalman Filter to estimate the tracking of images in new frames repeatedly. SORT uses constant speed and linear observations to determine the position of each existing path based on the location of coordinates before and after the frame changes [10].

III. System Design

Utilization of deep learning to detect vehicle objects in videos taken from drone. Video data obtained from drone in the form of aerial video contained vehicles with a certain speed. The work process of the system is shown in figure 1.

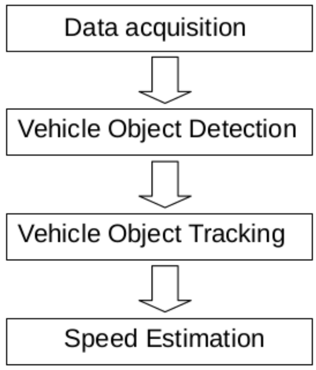

Fig. 1: System Diagram Blocks

Based on the block diagram in figure 1, the video data of a vehicle taken from a drone with a certain speed to adjust the speed of the image system with the original speed. Data that has been labeled in the form of aerial vehicle imagery and then trained to create a class aerial model to detect vehicles according to the label. Detection uses YOLOv3 object detection with bounding boxes that match the training model. The object image of the vehicle that has been detected will go through the object tracking stage using the Kalman filter and SORT (Simple Online and Realtime Tracking) by distinguishing each vehicle with an ID. When the vehicle object that has been detected will be tracked and then the vehicle object that has entered the RoI (Region of Interest) in the initial frame to the end frame then calculates the distance traveled and the time taken by the detected vehicle object, speed calculation can be done with distance data and time that has been obtained.

A. Training Phase

In the training phase, the dataset required for the detecting a top view of the vehicle in the video. We trained from 5400 vehicles from several images in the internet as many as 4800 iterations. The weight-training results are used to detect objects in the proposed system. An object is detected as a class of cars, trucks and buses if the detected area exceeds the threshold. In this study, The threshold was 80% of the vehicle top view.

B. Data acquisition

There are two locations used in data collection, namely: Citaringgul Street, Bogor and M.H. Thamrin Street, Bogor. Video data taken from a drone with a vertical camera downward that reaches the RoI region as seen in figure 2. The altitudes of the drones are 10 meters, 12 meters and 15 meters. We condition the tested car speed is 10 Km/h, 20 Km/h, 30 Km/h, and 40 Km/h during the daytime. RoI is the area used to estimate the speed with the distance and time obtained when the vehicle passes through the area, by counting frames when entering up to frames when exiting RoI. There are 3 RoI distances used in this video shooting, the first is the RoI distance that is used 5 meters, the second RoI distance used is 10 meters, and third the RoI distance used is 15 meters.

C. Vehicle Detection

Detection of vehicles that have been labeled and trained will be detected using YOLOv3. The standard input image normally used by YOLOv3 is 416×416 pixels that predicted by three scales 32, 16 and 8. Layer 13×13 is used to detect large objects, while layer layer 26×26 detects medium objects, and 52×52 detects small objects. In the detection process, Objects are identified in all video frames.

D. Vehicle Tracking

Objects that have been detected every frame, to distinguish each object that has been tracked, an ID is added to each object as shown figure 3 The detected object will be tracked

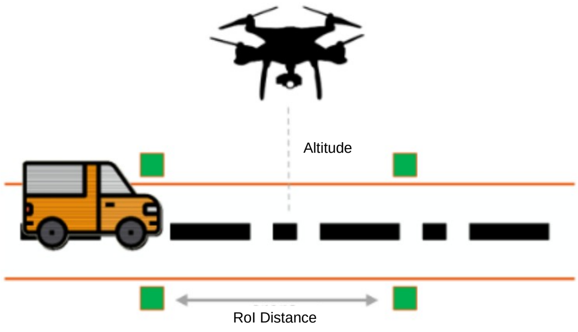

Fig. 2: RoI taken by Drone Camera

continuously. At this stage, the object is still in the next sequence frame in the video, the object will be given an id according to the previous frame.

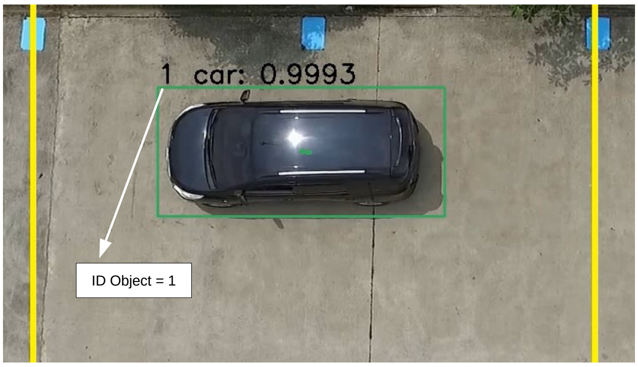

Fig. 3: ID on Object

E. Vehicle Speed

Calculation of the detected object speed starts when the object enters the RoI area from the starting point as in figure 4. When the frame has started video data will be calculated, as long as it passes through the RoI area the object will continue to move up to the last end point The frame calculation will stop as shown in figure 5. The number of frames that have been calculated will be divided by the fps of the video data entered to get the time value of the object passing through the RoI. After obtaining the time value of the object that passes through the RoI, then to get the speed of the object the time value will be used as a divisor of the distance RoI in video data. Obtained the value of the object's speed in m/s, to get the value speed of Km/h, the speed value of m/s multiplied by 3600/1000 following the equation 1.

IV. Experiments and Results

In this study, there are experiments that have been completed: Experiments Based on frame rate and altitude of drone. The resolution video is 1920 *times* 1080 pixels. The Frame Rate parameter will use 3 parameters namely: 15 fps, 25 fps and 30 fps. For altitude parameters drone using a height of 10 meters, 12 meters and 15 meters. RoI distance parameters will be entered with a distance of 5 meters, 10 meters and 15 meters.

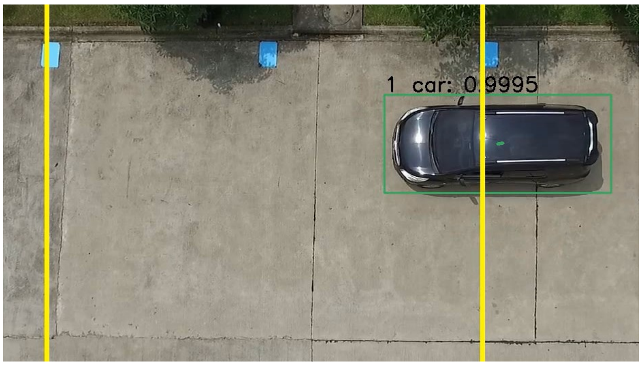

Fig. 4: Object Entering RoI

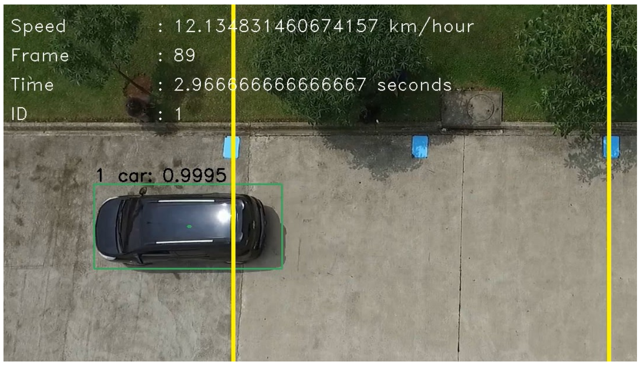

Fig. 5: Object Exit RoI

A. Experiment Results Based on Frame Rate

The results of the calculation of speed based on Frame Rate with RoI used 10 meters and the height of drone: 10 meters, 12 meters and 15 meters. The results can be seen in table I, table II, table III. From the plot in figure 6, the error will decrease if the video frame rate is increased. Increasing the number of frame rates causes increasing the number of images to be analyzed so that the accuracy of the speed calculation increases. However, Error of frame rate 25 fps has an error similar to the frame rate 30 fps.

TABLE I: Car speed error based on frame ate, drone altitude is 10 meters.

Frame Rate (fps)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
15	44,90%	55%	60,90%	43,75%	51,14%
25	8,40%	10%	18,60%	7,13%	11,03%
30	9,10%	10%	14,30%	11,50%	11,23%

TABLE II: Car speed error based on frame rate, drone altitude is 12 meters.

Frame Rate (fps)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
15	20,60%	41,30%	50%	48,08%	40,00%
25	55,20%	15,40%	0%	6,25%	19,21%
30	54,30%	14,90%	0%	3,58%	18,20%

TABLE III: Car speed error based on frame rate, drone altitude is 15 meters.

Frame Rate (fps)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
15	46,00%	35,70%	43,73%	43,75%	42,30%
25	7,10%	32,35%	15,40%	12,50%	16,84%
30	8%	31,70%	15,13%	12,50%	16,83%

B. Experiment Results Based on the Altitude of Drone

The results of the calculation of speed based on altitude and the length of ROI are 5 meters, 10 meters and 15 meters that can be seen at table IV, table V, table VI and figure 7. The error increases as the drone altitude increases because the object looks small. It causes vehicle misidentification on some frames. Also the higher the drone flies, the stronger the wind blowing towards the drone.

We used the Drone DJI Phantom 3 Professional, camera angle of 94 degrees, video resolution of 1920 $\times$ 1080 pixels. The video speed calculation is processed using a computer CPU Intel Core i7-8750H, GPU Nvidia GeForce 1050Ti, Memory 8 GB.

TABLE IV: Car speed error based on altitude, distance of ROI 5 meters.

High (m)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
10	5,35%	1,75%	7,13%	4,94%	4,79%
12	0,58%	13,90%	11,17%	12,14%	9,45%
15	4,42%	23,19%	6,01%	7,50%	10,28%

TABLE V: Car speed error based on altitude, distance of ROI 10 meters.

High (m)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
10	4,21%	3,02%	4,67%	5,60%	4,38%
12	0,72%	4,25%	3,35%	10%	4,58%
15	2,53%	3,3%	2,59%	10%	4,61%

TABLE VI: Car speed error based on altitude, distance of ROI 15 meters.

High (m)	Error				
	10 km/hour	20 km/hour	30 km/hour	40 km/hour	Average
10	1,75%	2,70%	0,58%	6,81%	2,96%
12	7,91%	2,39%	7,66%	1,81%	4,94%
15	1,80%	6,33%	6,41%	8,50%	5,76%

C. Experiment Results Based on the length of ROI

The results of the calculation of speed based on length of ROI and the height between drone and road are 10 meters, 12 meters and 15 meters is depicted in figure 8. The data of graphic are derived from table IV, table V, table VI. Figure 8, The experimental results show that the closer of ROI distance causes a large error. However, the anomaly occurred at an ROI distance of 15 meters at altitude 12 meters and 15 meters. The error of them is greater than the distance ROI 10 meters. This is probably due to the influence of wind blows.

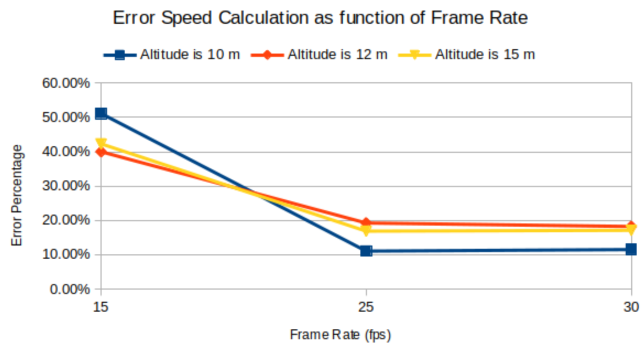

Fig. 6: Error Speed Calculation based on frame rate, altitude values are 10,12,15 meters.

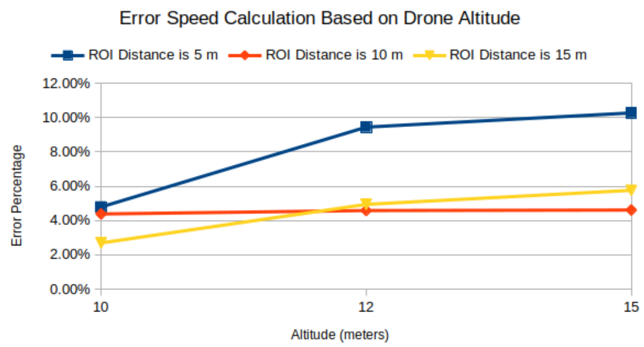

Fig. 7: Error Speed Calculation based on altitude, ROI distances are 5,10,15 meters.

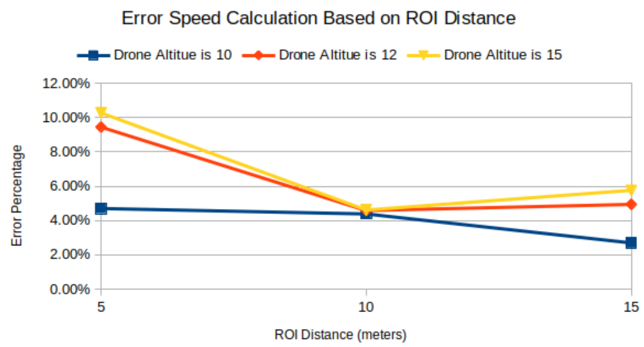

Fig. 8: Error Speed Calculation based on ROI distance, altitudes of drone are 10,12,15 meters.

V. Conclusion

In this study, we proposed method for calculation speed using Deep Learning using YOLO library. Our experiment obtained several conclusion, namely: Calculation of speed with difference testing frame rate with a height of 10 meters, 12 meters, and 15 meters. The average indicates the result of the error value decreases as the frame rate (fps) value increases. However there are some anomalies, for example in the table I fps 25 to 30 at speed of 40km/h, that is probably due to the wind factor. The estimated calculation of vehicle speed is influenced by the altitude of the drone. Because the higher the drone from the ground, the smaller the ROI area and the wind blowing to the drone is getting tighter. Since the longer the ROI causes the number of frames to increase, the speed calculation will be more accurate.

ACKNOWLEDGMENT

This research was supported by Computer Engineering Department, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] "Perkembangan Jumlah Kendaraan Bermotor Menurut Jenis 1949 - 2018," Badan Pusat Statistik, [Online]. Available: https://www.bps.go.id/dynamictable/2016/02/09%2000:00:00/1133/perkembangan-jumlah-kendaraan-bermotor-menurut-jenis-1949-2017.html. [Accessed: 20-Sep.-2020].

[2] A. Kurniawan, A. Ramadlan and E. M. Yuniarno, "Speed Monitoring for Multiple Vehicle Using Closed Circuit Television (CCTV) Camera," 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM 2018), Surabaya, Indonesia, pp. 88-93, 2018.

[3] S. B. Prakoso, "Pengembangan Sistem Penghitung Jumlah Kendaraan Memanfaatkan Citra Aerial yang Diambil dengan Pesawat Tanpa Awak," Undergraduate thesis, Departemen Teknik Komputer, Institut Teknologi Sepuluh Nopember, Surabaya, 2019.

[4] N. H. Tsani, "The implementattion of vehicle speed detection using webcam with frame difference method," bachelor thesis - Telkom University, August, 2017.

[5] A. M. Samad, N. Kamarulzaman, M. A. Hamdani, T. A. Mastor, and K. A. Hashim, "The potential of unmanned aerial vehicle (uav) for civilian and mapping application," in 2013 IEEE 3rd International Conference on System Engineering and Technology, pp. 24572461, August 2013.

[6] A. Shresta, A. Mahmood. "Review of Deep Learning Algorithms and Architectures," Department of Computer Science and Engineering, University of Bridgeport, 2019.

[7] J. Lu, C. Ma, L. Li, X. Xing, Y. Zhang, Z. Wang, and J. Xu. A vehicle detection method for aerial image based on yolo Journal of Computer and Communications, 2018, 6, 98-107, 2018.

[8] F. Miao, Y. Tian and L. Jin. Vehicle direction detection based on YOLOv3, School of Information Engineering, Communication University of China, 2019.

[9] J. Redmon, A. Farhadi. YOLOv3: An Incremental Improvement, arXiv:1804.02767, 2018.

[10] X. Hou, Y. Wang, and L. Chau. "Vehicle Tracking Using Deep SORT with Low Confidence Track Filtering," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 2019, pp. 1-6.

[11] A. B. Raharjo, M. Quafafou. Penggabungan Keputusan Pada Klasifikasi Multi-label, Jurnal Ilmiah Teknologi Informasi - Volume 13, Nomer 1, Januari 2015.