Vehicle pose estimation for vehicle detection and tracking based on road direction

Adhi Prahara [1,*], Ahmad Azhari [2], Murinto [3]

Informatics Department, Universitas Ahmad Dahlan, Umbul Harjo, Yogyakarta 55164, Indonesia
[1] *adhi.prahara@tif.uad.ac.id*; [2] ahmad.azhari@tif.uad.ac.id; [3] murintokusno@tif.uad.ac.id*
* corresponding author

ARTICLE INFO

ABSTRACT

Vehicle has several types and each of them has different color, size, and shape. The appearance of vehicle also changes if viewed from different viewpoint of traffic surveillance camera. This situation can create many possibilities of vehicle poses. However, the one in common, vehicle pose usually follows road direction. Therefore, this research proposes a method to estimate the pose of vehicle for vehicle detection and tracking based on road direction. Vehicle training data are generated from 3D vehicle models in four-pair orientation categories. Histogram of Oriented Gradients (HOG) and Linear-Support Vector Machine (Linear-SVM) are used to build vehicle detectors from the data. Road area is extracted from traffic surveillance image to localize the detection area. The pose of vehicle which estimated based on road direction will be used to select a suitable vehicle detector for vehicle detection process. To obtain the final vehicle object, vehicle line checking method is applied to the vehicle detection result. Finally, vehicle tracking is performed to give label on each vehicle. The test conducted on various viewpoints of traffic surveillance camera shows that the method effectively detects and tracks vehicle by estimating the pose of vehicle. Performance evaluation of the proposed method shows 0.9170 of accuracy and 0.9161 of balance accuracy (BAC).

I. Introduction

Vehicle detection and tracking are one of the important stages in Intelligent Traffic Surveillance System (ITSS). The system has purposes to give reliable traffic information for traffic management and control. In order to give reliable traffic information, the system should be robust to problems, which found in real traffic situation such as occlusion, variety of vehicle types, and variety viewpoints of traffic surveillance camera. Occlusion is commonly caused by traffic congestion. This condition becomes challenge in vehicle detection because the vehicle is partially covered with another vehicle or another object on the road which makes only part of vehicle is visible from the video camera [1][2]. This condition commonly occurs in the front or rear viewpoint of traffic surveillance camera.

The other problems come from a large variety of vehicles with cluttered environment, various illumination, and driving behavior [3]. Modern vehicle has many types and each of them has different color, size, and shape. Even appearance of a vehicle looks different if viewed from various viewpoints of traffic surveillance camera resulting in many possibilities of vehicle orientation i.e. oncoming, preceding, or side view [4]. Not only vehicle orientation, the configuration of traffic surveillance camera also greatly affects the lighting condition and output quality of traffic image or video [5]. This problem might be handled by installing traffic video cameras in a fixed position to give the fixed orientation of viewpoint. However, in most countries, many traffic surveillance cameras that already installed on the road are not configured for intelligent traffic surveillance purposes. Rather than fix the position of traffic surveillance camera, which might be costly, the system itself should adapt to the current situation.

Many methods of vehicle detection and tracking require pre-defined detection area and vehicle orientation [6]. It is usually done manually by selecting region of interest, which is road region or road lanes, and determines the vehicle orientation. The road region is useful to localize vehicle detection area. Vehicle orientation is used to determine the vehicle detectors used by the system. However, the manually predefined method is not efficient if there are many traffic surveillance cameras to be configured.

In order to detect and to track vehicle in multi-view orientations, vehicle detection method can utilize vehicle models in various orientations [7][8] or using part-based vehicle models [9][10]. The first method usually employs real vehicle data in various orientations or 3D vehicle models to generate vehicle poses in multi orientations and builds vehicle detectors for each vehicle orientation. The vehicle detectors will be chosen according to the traffic video situation. Vehicle detection that uses part-based vehicle models are designed to be robust in occlusion situation which only part of the vehicle is visible in traffic video.

Vehicle detection and tracking have been developed by many researches. Elkerdawi *et al.* proposed a real-time vehicle detection and tracking framework [6]. The vehicle was detected using Haar-like features and Adaboost cascade classifier. The detected vehicle was associated with a tracking method, namely Compressive Tracking (CT), to increase its detection accuracy and removed repetitive detection of the same vehicle over frames. The result showed robust performance under different illumination, heavy flow, and multiple target.

Tian *et al.* proposed a method to detect and to track rear viewed vehicle based on multiple vehicle parts [11]. The parts such as license plate and rear lamps were extracted its color, texture, and region feature then treated as graph nodes to construct probabilities graph using Markov Random Field model. Loopy belief propagation was used in the vehicle detection process and Kalman filter was used in the vehicle tracking process. The method showed good result in partial occlusion, various lighting conditions, and achieved real time performance.

Cao *et al.* in their research proposed a method to detect and to analyze vehicle on low attitude traffic video [12]. Video data were extracted using Light-Boost Pyramid Sampling Histogram of Oriented Gradients (LBPS-HOG), which was a variant of HOG. The features were classified using Support Vector Machine (SVM) then matching the color and spatial features of vehicle from frame to frame to track the vehicle. The experiment showed better result than other methods in the same dataset.

Jazayeri *et al.* proposed vehicle detection and tracking based on vehicle movement [13]. They used Hidden Markov Model (HMM) to differentiate the vehicle with its background then performed probabilistic tracking. Corner features, horizontal line segment, and light intensities combined with vehicle behavior were used to track vehicle ID. The result showed good accuracy with 86.6% of true positive and 13.2% of false positive.

Kovačić *et al.* proposed a method to detect and to track vehicle in multiple lanes on the road [14]. They used the background and foreground segmentation to detect moving vehicle. Every cluster from blob object was assumed as foreground then they were selected and grouped into vehicle or non-vehicle category. Vehicle tracking was done using vehicle cluster analysis, such as grouping neighborhood pixel on foreground object and performing weighting based on cluster area, overlapping cluster area and distance between clusters. For real time detection and tracking, CPU and GPU were employed. The accuracy achieved in their research up to 95%.

This research proposes a method to estimate the pose of vehicle for vehicle detection and tracking based on road direction. In summary, the contribution of this work is given as follows:
a. The proposed method uses 3D vehicle models to represent vehicle orientation in various viewpoints of traffic surveillance camera.
b. The vehicle detectors are generated from 3D vehicle models which grouped into four-pairs orientation of viewpoint to handle two-ways road and reduce the number of detectors.
c. Road area is extracted on traffic surveillance image to localize the detection area and road direction is calculated to estimate the pose of vehicle.
d. The result of vehicle pose estimation is used to select a suitable vehicle detector for vehicle detection process.
e. The method uses Histogram of Oriented Gradients (HOG) and Linear-SVM (Linear-Support Vector Machine) in multi-scale detection window to perform vehicle detection.

f. The final vehicle object is obtained after applying vehicle line checking method on vehicle detection result.

g. Finally, vehicle tracking is performed to label the vehicle.

The rest of this paper is organized as follows: Section II presents the proposed method. Section III presents the result and discussion. Finally, the conclusion of this work is described in Section IV.

II. Methodology

The general steps of vehicle pose estimation for vehicle detection and tracking based on road direction method is shown in Fig. 1. From Fig. 1, the proposed method begins with training vehicle detectors in four pair orientation categories then extracts the road area and estimates the pose of vehicle based on road direction. The pose of vehicle is used to select vehicle detector for vehicle detection process then obtains the final vehicle object by applying vehicle line checking method on vehicle detection result. Finally, perform vehicle tracking to label the vehicle.
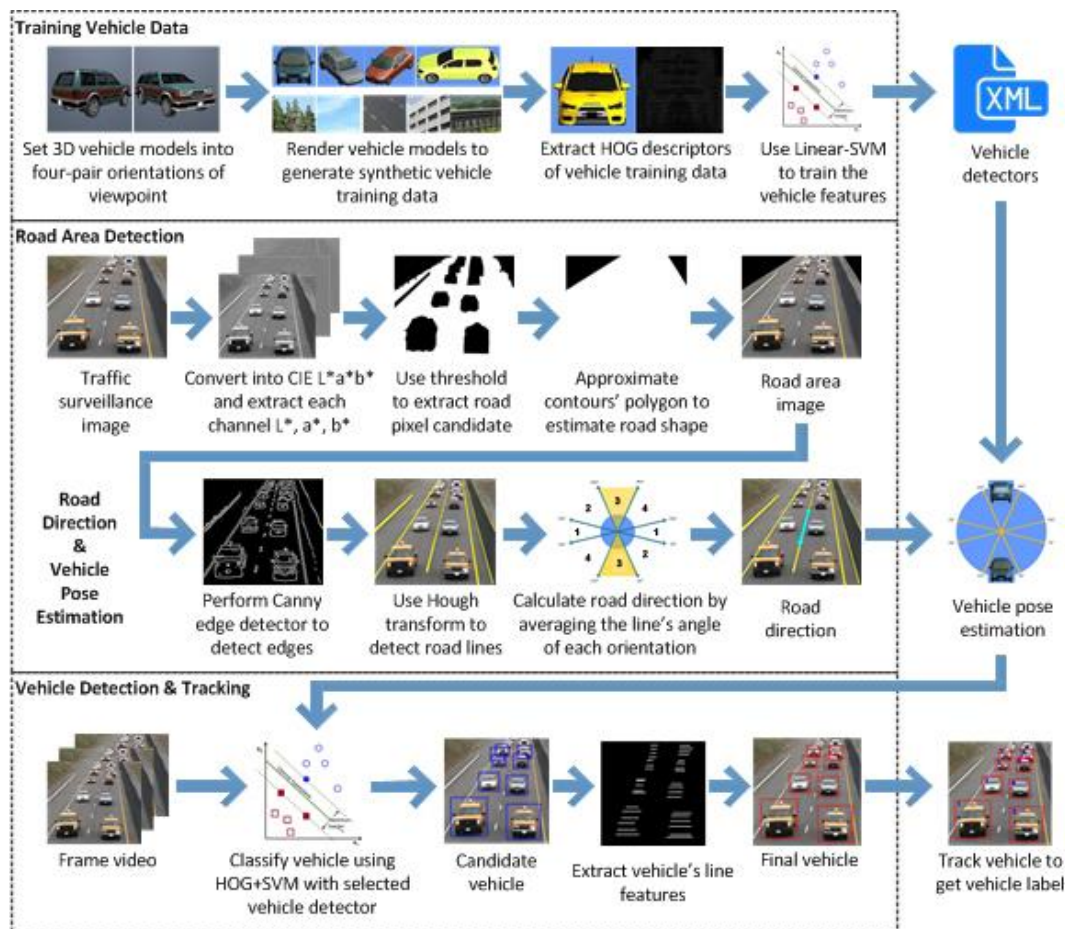

Fig. 1. The general steps of vehicle pose estimation for vehicle detection and tracking based on road direction method

A. Dataset

The training data consists of positive vehicle images which generated from 3D vehicle models and negative data which consist of non-vehicle objects as shown in Fig. 2a and Fig. 2b respectively. This research utilizes 3D vehicle models because of its effectiveness in modelling vehicle pose in any orientation [7][9]. From Fig. 2a, vehicle positive data consist of various types of vehicle models such as car, pickup, van, bus, and truck which generated in various orientations of viewpoint. From Fig. 2b, the negative data are trees, sky, road, buildings, and road signs. The test data are traffic surveillance image and video in various orientations of viewpoint.
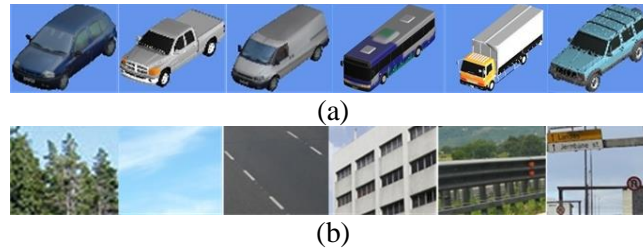
(a)

(b)

Fig. 2. Vehicle training data (a) Positive training data (b) Negative training data

B. Road Area Extraction

Road can be extracted using road classifier on road features [15][16] or using statistic of road color information [17]. In this research, road color threshold and road texture information are used to extract the road pixels, then morphological operation is applied to enhance the result. The procedure to extract road pixel on traffic surveillance image is explained below.

1. Traffic surveillance image is converted into CIE L*a*b* color space.
2. Road pixel candidate is extracted on every channel of CIE L*a*b* color space using (1) for L* channel, (2) for a* channel, and (3) for b* channel.

$$road_L(i,j) = \begin{cases} 0.2 \leq L(i,j) \leq 0.8, 1 \\ otherwise, \quad\quad 0 \end{cases} \tag{1}$$

$$road_a(i,j) = \begin{cases} 0.48 \leq a(i,j) \leq 0.53, 1 \\ otherwise, \quad\quad\quad 0 \end{cases} \tag{2}$$

$$road_b(i,j) = \begin{cases} 0.48 \leq b(i,j) \leq 0.53, 1 \\ otherwise, \quad\quad\quad 0 \end{cases} \tag{3}$$

3. Road pixel candidate is calculated using (4).

$$road = road_L \cdot road_a \cdot road_b \tag{4}$$

4. The homogeneity of road pixel candidate around its neighborhood is calculated to extract road texture information using (5) where $p(i,j)$ is the value on i,j GLCM (Gray Level Co-Occurrence Matrix) coordinate [18].

$$\sum_{i,j} \frac{p(i,j)}{1 + |i-j|} \tag{5}$$

5. The result of road pixel extraction is enhanced by applying morphological opening operation.
6. Road area is selected on road contours that larger than area threshold.
7. Finally, road shape is estimated by calculating convex hull using polygon approximation on road contours.

Color threshold is applied on CIE L*a*b* color channel because from various types of road samples, the chromatics channels a* and b* have small range of variance while only the luminance channel L* has large variance. Road texture is important to differentiate similar color of non-road pixel. Road texture has high homogeneity value because of its uniform texture.

C. Pose Estimation of Vehicle Based on Road Direction

Pose of vehicle usually follows road direction. By calculating road direction, pose of vehicle can be estimated. Road direction is calculated from the existence of road lines, road lanes, and road shapes. The procedure to calculate road direction is explained below.

1. Road edges are extracted from road area using Canny edge detector with only strong edges are preserved.
2. The edges will be used to estimate road lines and road lanes using Hough transform.
3. The angle of road lines and road lanes are grouped into four-pair orientations category, $orientation = 1 ... 4$ using (6) in clockwise orientation as shown in Fig. 3a.

$$orientation = \begin{cases} 60° \leq angle \leq 120°, & 3 \\ 240° \leq angle \leq 300°, & 3 \\ 340° \leq angle \leq 20°, & 1 \\ 160° \leq angle \leq 200°, & 1 \\ 120° < angle < 160°, & 4 \\ 300° < angle < 340°, & 4 \\ 20° < angle < 60°, & 2 \\ 200° < angle < 240°, & 2 \end{cases} \tag{6}$$

4. The average angle of each orientation category ($orientation\ angle(i)$) is calculated using (7) where $line\ angle(j)$ is the jth road line angle and N_i is the number of line angles in the ith orientation.

$$orientation\ angle(i) = \frac{\sum_{j=1}^{N_j} line\ angle(j)}{N_i} \tag{7}$$

5. The estimated road direction angle ($road\ angle$) is calculated using (8) where N is the number of orientation category which have at least one member.

$$road\ angle = \frac{\sum_{i=1}^{N} orientation\ angle(i)}{N} \tag{8}$$

6. If no road lines or road lanes found, the road direction angle is calculated by estimating the angle of minimum rectangle area from road shape.

Road lines and road lanes are the key features to calculate road direction [19]. These features can be extracted using Canny edge detector and Hough transform. Any possibility of road direction is grouped into four pair orientation categories. The pairing of orientation has purposes to deal with two-ways road.

In this research, the pose of vehicle is grouped into four pair orientations of viewpoint according to the road direction. The four orientations of viewpoint are front / back view, left / right side view, top right / bottom left view, and top left / bottom right view as shown in Fig. 3b. The pairing of viewpoint is proposed because the size, shape, and appearance of vehicle are similar in the same pairing orientation categories. Pose of vehicle in the same orientation category will be varied to cover every angle in the categories [20]. This grouping method can also significantly reduce the number of vehicle detectors.

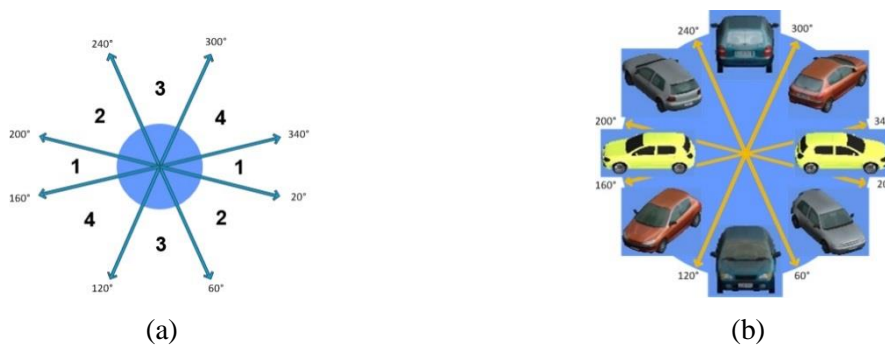(a)   (b)

Fig. 3. Pose estimation of vehicle (a) Four pair orientation categories to group the road line's angles (b) Four pair orientation categories of vehicle pose.

D. Vehicle Detection and Tracking

Vehicle can be differentiated from its background by extracting vehicle features and uses machine learning to classify the features. In this research, HOG is used as feature extraction method and SVM is used as classifier. The procedure of applying HOG descriptors on image is described below [21]:
1. Gamma and color from input image are normalized.
2. The gradient is calculated on normalized image.
3. For color image, gradients are computed for each color channel then take the one with the largest norm as the pixel's gradient vector.
4. Image is divided into small spatial regions called cells.

5. Accumulating a local 1-D histogram of gradient directions or edge orientations over the pixel inside the cell.
6. Orientation bins are evenly spaced over 0-180 degree (unsigned gradient) with 20 degrees for each space.
7. The nine orientation bins are filled with the weighted magnitude for each orientation.
8. The cells are organized into overlapping blocks.
9. Perform contrast normalization on the overlapping blocks.
10. HOG descriptors are obtained by collecting features on all blocks.

SVM is a classifier that determined by separation called hyperplane. The decision function of SVM is calculated using (9) [22].

$$sgn(w^T \phi(x) + b) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \right) \tag{9}$$

Where w is the solution vector, x_i is the input features, C is the regularization parameter, $\phi(x)$ maps x into a higher-dimensional space, b is bias, y_i is the label, α_i is a vector of weights, and $K(x_i, x)$ is the kernel function. The parameter $y_i \alpha_i \forall i,$ b, label names, support vectors, and kernel parameter are saved as an output trained model of SVM.

In this research, the vehicle detection process uses HOG and SVM which proven to produce good results in classifying vehicle [23][24][25]. The result of vehicle detection is a candidate vehicle. To ensure the detection, vehicle need to be checked its line features which exist in every pose of vehicle. Vehicle line features can be obtained from vehicle by applying Canny edge detector and estimating the line using Hough transform. After the final vehicle object is obtained, it needs to be labelled which useful to track individual vehicle and identify vehicle behavior. In this research, vehicle is tracked using optical flow [26]. The procedure to detect and to track vehicle is explained below:
1. Vehicle synthetic data which generated by 3D vehicle models in various orientation are grouped into four pair orientations of viewpoint.
2. The features of each vehicle orientation data are extracted using HOG and trained using Linear-SVM with *k*-fold cross validation. The cross validation method is used to determine the best parameter of SVM. The result of training is vehicle detectors of each orientation category.
3. The detection utilizes multi-scale window along with HOG and Linear-SVM to classify each window into candidate vehicle or background.
4. The vehicle line features from candidate vehicle is extracted using Canny edge detector and Hough transform. Candidate vehicle must pass the vehicle line threshold to be considered as final vehicle object.
5. Final vehicle object will be tracked using optical flow to give individual vehicle label.

E. Performance Evaluation

The evaluation procedure uses confusion matrix then calculates the accuracy, specificity, sensitivity, and precision using (10), (11), (12), and (13) respectively where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. BAC (Balance Accuracy) which calculated using (14) is also used for evaluation because of the imbalance between positive and negative test data [27].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{11}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{12}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{13}$$

$$BAC = \frac{Specificity + Sensitivity}{2} \tag{14}$$

III. Result and Discussion

The proposed method is built using C++ language with additional OpenCV library [28] and runs on notebook Intel Core i5, 8 GB of RAM, and NVidia GT540M 2GB. The training data consist of 2500 positive images and 6000 negative images with size 128x64 for left / right orientation and 64x64 for other orientation categories. Vehicle detectors are trained using Linear-SVM with 10-fold cross validation procedure. The performance of the proposed method is tested on 350 traffic surveillance images for vehicle detection and 4 traffic surveillance video for vehicle tracking in various road directions.

A. Road Area Extraction Result

Road area extraction is used to localize the detection area. The localization is important to remove non-road area, which likely do not contain vehicle object. The step by step results of road extraction are shown in Fig. 4. Fig. 4a shows a frame from traffic surveillance video. Fig. 4b shows the results of road pixel extraction by thresholding the road color and homogeneity feature. Fig. 4c shows the results of road shape estimation by applying morphological opening operation and calculates the convex hull of minimum area of rectangle from road contours. Fig. 4d shows the results of road detection on frame which marked by magenta color marker.

(a)   (b)   (c)   (d)

Fig. 4. The result of road area extraction (a) Road frame (b) Road pixel extraction result (c) Road shape estimation result (d) Road area extraction on frame (magenta color marker).

As shown in Fig. 4c, the small blob area in Fig. 4b is removed because they are lower than area threshold which is 0.05 of image size. The road area is assumed large because the traffic surveillance camera will point to the road to cover wide view of traffic. In Fig. 4b, the road shape is well estimated by calculating the convex hull of polygon from road contours. However, sometimes the non-road area is also recognized as road area by the polygon's fitting process as shown in Fig. 5. In Fig. 5a, both of the sidewalks are recognized as road area and in Fig. 5b, the wall on the right side of the road is also recognized as road area.
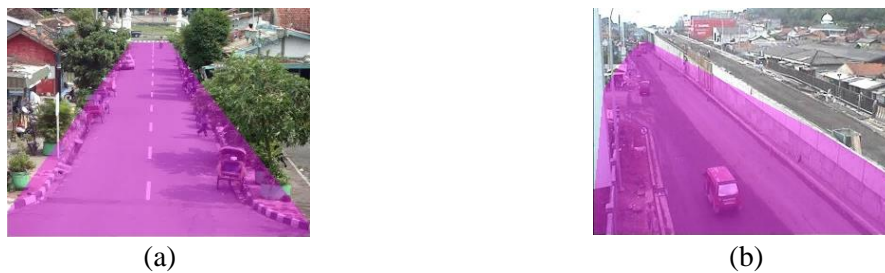

(a)   (b)

Fig. 5. Non-road area which detected as road area (a) Both of the sidewalks are detected as road area (b) The wall on the right side of the road is detected as road area

B. Vehicle Pose Estimation Result

Pose of vehicle is estimated from road direction while the road direction is calculated by averaging the angles of road lines, road lanes, or road shape in the four bins of orientation categories. The result of vehicle pose estimation is shown in Fig. 6 by index number which represents the four orientation categories of vehicle pose which described in Fig. 3, magenta color marker which represents road area, and arrow marker with cyan color which represents road direction.

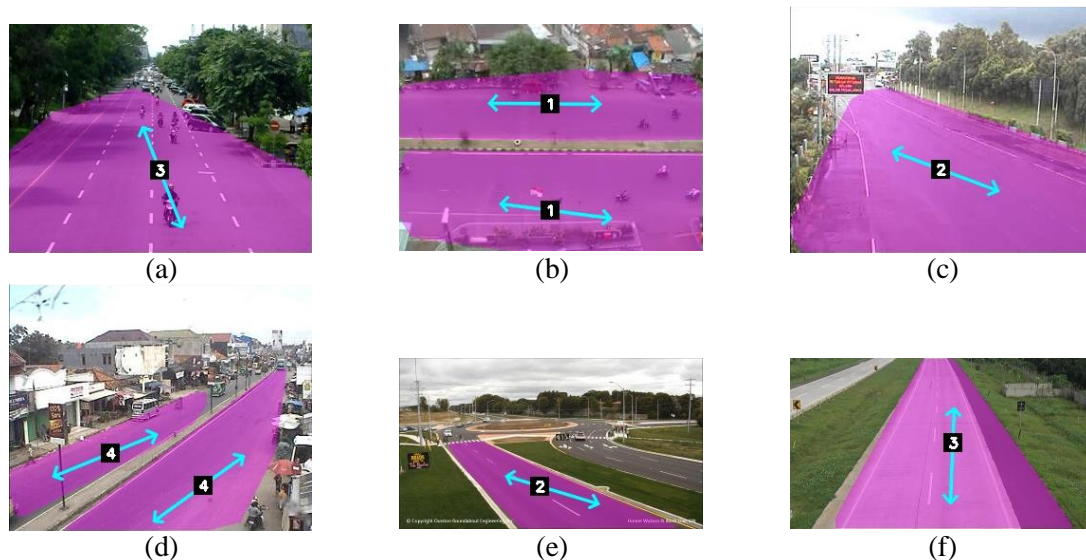| (a) | (b) | (c) |
| (d) | (e) | (f) |

Fig. 6. Result of vehicle's pose and road direction estimation (a) Front / back view, (b) Left side / right side view, (c) Top left / bottom right view, (d) Top right / bottom left view, (e) Top left / bottom right view, and (f) Front / back view of vehicle's pose.

From Fig. 6, Fig. 6a and Fig. 6f show the result of vehicle's pose estimation as front / back orientation, Fig. 6b shows the result as left / right side orientation, Fig. 6c and Fig. 6e show the result as top left / bottom right orientation and Fig. 6d shows the result as top right / bottom left orientation. As shown in Fig. 6, the proposed method shows good result in calculating road direction and estimates the pose of vehicle.

However, the method is only effective for straight road type and cannot handle complex road shapes such as intersection, roundabout, or winding road. Straight road type is preferable for Intelligent Traffic Surveillance System (ITSS) than the other road types, especially for vehicle counting and vehicle tracking because there is only one or two road direction. The method is also effective if road lines, road lanes, and road shape are found on the road image because the road direction and the pose estimation of vehicle are based on those road features.

The result of vehicle pose estimation is used to determine the suitable vehicle detector in the detection process. To cover various poses of vehicle, a number of vehicle detectors or single vehicle detector with large variation of training data is needed. By using only one suitable vehicle detector in selected orientation category, the number of vehicle detectors used in the detection process can be reduced and the training data used in the training process can be minimized.

C. Vehicle Detection Result

The vehicle detection uses HOG and Linear-SVM with vehicle detector that selected according to the result of vehicle's pose estimation. Multi-scale detection window with scale factor 1.02 and scale level 64 is used to handle a variation of vehicle size when they are near or far from the traffic surveillance camera. The detection threshold is 5 as a requirement for the detected area to be considered as vehicle object. Apart from the vehicle detection procedure, the vehicle object also must pass the vehicle line threshold in order to be considered as final vehicle object.

The result of vehicle detection is shown in Fig. 7 by red rectangle marker. From Fig. 7, Fig. 7a and Fig. 7f show the result of vehicle detection in front / back orientation, Fig. 8b in left / right side orientation, Fig. 8c and Fig. 8e in top left / bottom right orientation, and Fig. 8d in the top right / bottom left orientation. As shown in Fig. 8, the proposed method can detect vehicles in various orientations by selecting a suitable vehicle detector that matches the pose of vehicle.

Sometimes, the proposed method fails to detect vehicle because the vehicle is too near or too far from the traffic surveillance camera, the vehicle is partially covered with the other object, or the vehicle is not following road direction. Fig. 8 shows the failure detection cases in vehicle detection process. Fig. 8a shows the small vehicle at a distance which failed to be detected because the vehicle is too far from the camera. Still in Fig. 8a, the two black vehicle that overlap with each other is also failed to be detected because the vehicle is partially covered with each other. Fig. 8b shows the failure

detection of vehicle which parked on the top side of the road because it is too far from the camera and on the bottom side of the road because it not follows road direction.
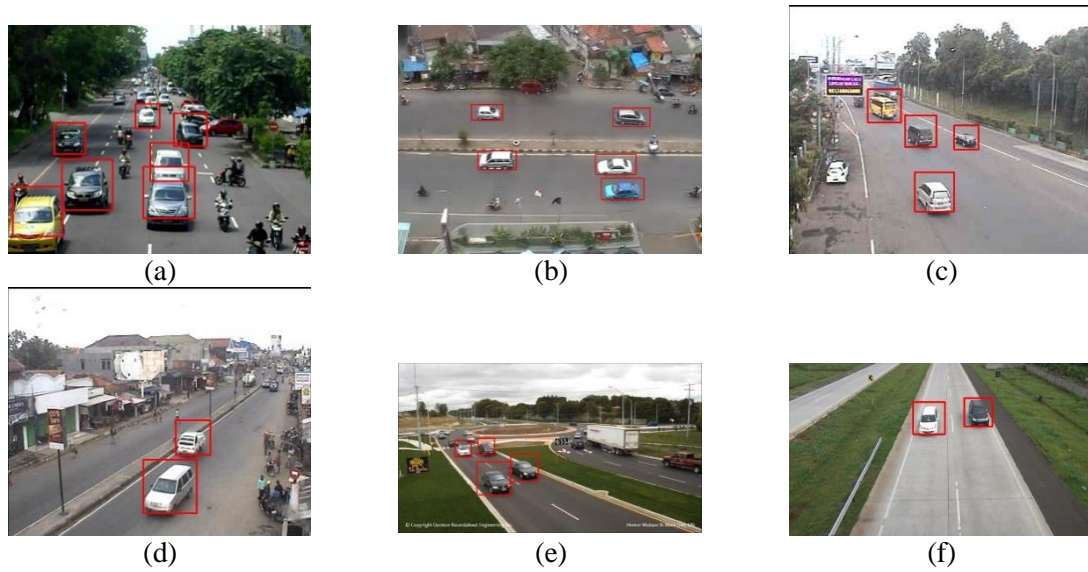

Fig. 7. Vehicle detection result (red rectangle marker). (a) Front / back orientation. (b) Left / right orientation. (c) Top left / bottom right orientation. (d) Top right / bottom left orientation. (e) Top left / bottom right orientation. (f) Front / back orientation.
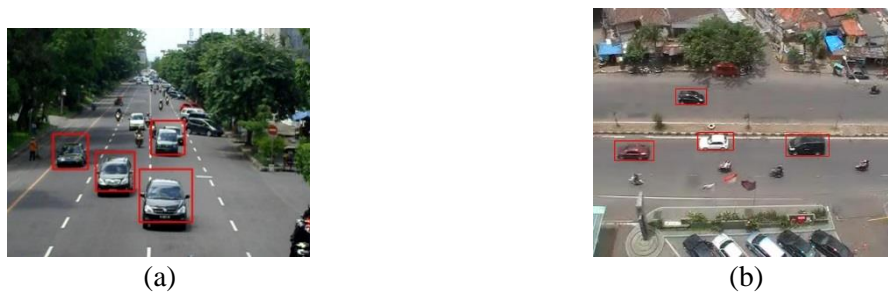

Fig. 8. Failure detection cases in vehicle detection process. (a) The case of vehicle which too near or too far from the camera and when partially covered by another object. (b) The case of vehicle that not follows road direction.

D. Vehicle Tracking Result

The vehicle tracking process uses optical flow to track the position of vehicle in the next frame and to label the same vehicle. The current position of the vehicle is translated into the next position by matching 20 features from the vehicle. This new translated position is compared with the detected vehicle on the next frame to identify whether it is the same vehicle that belongs to the previous detection or a new vehicle. Fig. 9 shows the result of vehicle tracking in traffic surveillance video.
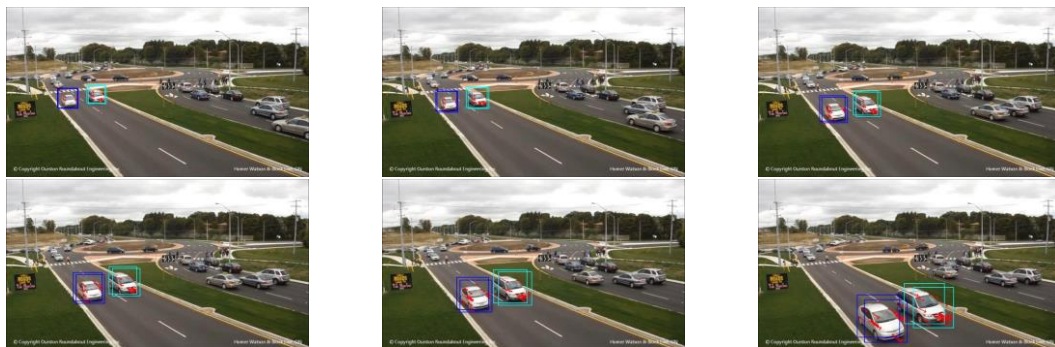

Fig. 9. Vehicle tracking result on traffic video

From Fig. 9, the detected vehicle is marked with two rectangles which represents the current vehicle position and predicted position on the next frame, red arrow marker represents the direction and distance between the current vehicle position with the next position on the next frame, and colored rectangles which label the same vehicle with the same color. Fig. 9 shows that the tracking method can track vehicle until it leaves the frame and labels individual vehicle correctly.

E. Performance Evaluation

The performance of vehicle's poses estimation for vehicle detection and tracking based on road direction method is evaluated using confusion matrix as shown in Table 1 and measured its accuracy, sensitivity, specificity, precision, and balance accuracy as shown in Table 2. In Table 1 and Table 2, the proposed method is accurate in detecting vehicle due to its high precision score, which is 0.9782. The high precision is mainly caused by the effectiveness of 3D vehicle models that can represent vehicle in any viewpoint and the pose estimation of vehicle that estimates the right pose of the vehicle and uses the right vehicle detector in the detection process.

Despite of its effectiveness, 3D vehicle model is expensive and difficult to make. The number of 3D vehicle model used in this research is 80 models. The limited number of 3D vehicle models used in the training process mainly causes the less hit rate shown by 0.8506 of sensitivity. However, the overall performance of the proposed method is good with 0.9170 of accuracy and 0.9161 of balance accuracy (BAC). BAC is also used for measurement because of imbalance between positive and negative test data.

Table 1. Confusion matrix of the proposed method

		Predicted Classification	
		Vehicle	*Background*
Actual Classification	*Vehicle*	809	18
	Background	142	960

Table 2. Performance evaluation of proposed method

Accuracy	Sensitivity	Specificity	Precision	BAC
0.9170	0.8506	0.9815	0.9782	0.9161

The proposed method chooses one suitable vehicle detector from four vehicle detectors based on the result of vehicle's pose estimation. This procedure has purposes to reduce the number of vehicle detector used in the detection process by choosing vehicle detector that matches the pose of vehicle. To measure its effectiveness, the proposed method is compared with the same vehicle detection method, but uses all vehicle detectors in its detection process. Table 3 and Table 4 show the confusion matrix and performance evaluation of vehicle detection that uses all vehicle detectors.

Table 3. Confusion matrix of vehicle detection using all vehicle detectors

		Predicted Classification	
		Vehicle	*Background*
Actual Classification	*Vehicle*	873	224
	Background	78	754

Table 4. Performance evaluation of vehicle detection using all vehicle detectors

Accuracy	Sensitivity	Specificity	Precision	BAC
0.8434	0.9179	0.7709	0.7958	0.8444

From Table 2 and Table 4, the performance scores of proposed method that uses one suitable vehicle detector is higher than the method that uses all vehicle detectors. Although the sensitivity scores of vehicle detection method that used all vehicle detectors is higher, but many false positives are also generated by this method. The low specificity score mainly caused by many false positives (background that detected as vehicle) as shown in Table 3. High hit rate is good but the precision also should be considered especially for vehicle tracking. By using one suitable vehicle detector which matches the pose of vehicle, the proposed method can reduce the false positive (high specificity score) and increases the true positive (high precision score).

The result of vehicle detection by multi-scale detection will be checked by vehicle line checking method before it becomes final vehicle object. To measure the effectiveness of vehicle line checking, the performance of proposed method with and without vehicle line checking is compared. Table 5 and Table 6 show the confusion matrix and performance evaluation of proposed method without vehicle line checking respectively.

As shown in Table 2 and Table 6, the performance of proposed method without vehicle line checking is lower than the one with vehicle line checking, especially the specificity value. As shown in Table 5, the number of background that detected as vehicle is high. As shown in Table 1, by checking the vehicle line after the detection process, the number of false positive is reduced and although the sensitivity slightly goes lower but the other performances is risen.

Table 5. Confusion matrix of proposed method without vehicle line checking

		Predicted Classification	
		Vehicle	*Background*
Actual Classification	*Vehicle*	821	196
	Background	130	782

Table 6. Performance evaluation of proposed method without vehicle line checking

Accuracy	Sensitivity	Specificity	Precision	BAC
0.8310	0.8633	0.7995	0.8072	0.8314

IV. Conclusion

This paper has discussed a method to estimate pose of vehicle for vehicle detection and tracking based on road direction. The proposed method has obtained good result in detecting and tracking vehicle by estimating pose of vehicle to determine the suitable vehicle detector used in the detection process. This research shows that vehicle training data can be represented effectively by using 3D vehicle models. Road area extraction is important to localize the detection area. Pose of vehicle can be estimated from road direction. HOG and SVM are good to be used for vehicle detection even with only a few number synthetic training data to build the detectors. The line features in every pose of vehicle is effective to reduce the false positive generated by vehicle detector. The proposed method also shows that by using one suitable vehicle detector that matches the pose of vehicle can perform better than firing all vehicle detectors.

The further development of this method is expected to be able to detect all kind of vehicles not only four wheels but also two wheels' vehicle on any viewpoint from traffic surveillance camera. The detection and tracking is not only effective for straight road type but also for roundabout, winding road, and intersection. The method is also expected to handle occlusion and to achieve better performance in vehicle detection.

References

[1] Y. Li, B. Li, B. Tian, and Q. Yao, "Vehicle Detection Based on the and– or Graph for Congested Traffic Conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 984–993, Jun. 2013.

[2] H. Van Pham and B.-R. Lee, "Front-view car detection and counting with occlusion in dense traffic flow," *Int. J. Control. Autom. Syst.*, vol. 13, no. 5, pp. 1150–1160, Oct. 2015.

[3] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle Detection Techniques for Collision Avoidance Systems: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2318–2338, Oct. 2015.

[4] S. Sivaraman and M. M. Trivedi, "Vehicle Detection by Independent Parts for Urban Driver Assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1597–1608, Dec. 2013.

[5] H. He, Z. Shao, and J. Tan, "Recognition of Car Makes and Models From a Single Traffic-Camera Image," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3182–3192, Dec. 2015.

[6] S. M. Elkerdawi, R. Sayed, and M. ElHelw, "Real-Time Vehicle Detection and Tracking Using Haar-Like Features and Compressive Tracking," Springer International Publishing, 2014, pp. 381–390.

[7] K.-H. Lee, J.-N. Hwang, and S.-I. Chen, "Model-Based Vehicle Localization Based on 3-D Constrained Multiple-Kernel Tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 38–50, Jan. 2015.

[8] Y. Tang, C. Zhang, R. Gu, P. Li, and B. Yang, "Vehicle detection and recognition for intelligent traffic surveillance system," *Multimed. Tools Appl.*, pp. 1–16, Mar. 2015.

[9] T. Wu, B. Li, and S.-C. Zhu, "Learning And-Or Models to Represent Context and Occlusion for Car Detection and Viewpoint Estimation," Jan. 2015.

[10] C. Wang, Y. Fang, H. Zhao, C. Guo, S. Mita, and H. Zha, "Probabilistic Inference for Occluded and Multiview On-road Vehicle Detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 215–229, Jan. 2016.

[11] Bin Tian, Ye Li, Bo Li, and Ding Wen, "Rear-View Vehicle Detection and Tracking by Combining Multiple Parts for Complex Urban Surveillance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 597–606, Apr. 2014.

[12] X. Cao, C. Wu, J. Lan, P. Yan, and X. Li, "Vehicle Detection and Motion Analysis in Low-Altitude Airborne Video Under Urban Environment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1522–1533, Oct. 2011.

[13] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle Detection and Tracking in Car Video Based on Motion Model," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 583–595, Jun. 2011.

[14] K. Kovačić, E. Ivanjko, and H. Gold, "Real time vehicle detection and tracking on multiple lanes," 2014.

[15] J. M. Á. Alvarez and A. M. Lopez, "Road Detection Based on Illuminant Invariance," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 184–193, Mar. 2011.

[16] J. Fritsch, T. Kuhnl, and F. Kummert, "Monocular Road Terrain Detection by Combining Visual and Spatial Information," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1586–1596, Aug. 2014.

[17] Y. He, H. Wang, and B. Zhang, "Color-Based Road Detection in Urban Traffic Scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[18] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[19] X. Du and K. K. Tan, "Vision-based approach towards lane line detection and vehicle localization," *Mach. Vis. Appl.*, vol. 27, no. 2, pp. 175–191, Feb. 2016.

[20] L. C. León and R. Hirata, "Car detection in sequences of images of urban environments using mixture of deformable part models," 2014.

[21] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893.

[22] C.-C. Chang and C.-J. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[23] A. Prahara and Murinto, "Car detection based on road direction on traffic surveillance image," in *2016 2nd International Conference on Science in Information Technology (ICSITech)*, 2016, pp. 344–349.

[24] T. Moranduzzo and F. Melgani, "Detecting Cars in UAV Images With a Catalog-Based Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[25] S. Madhogaria, P. Baggenstoss, M. Schikora, W. Koch, and D. Cremers, "Car detection by fusion of HOG and causal MRF," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 575–590, Jan. 2015.

[26] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann Publishers Inc., pp. 674–679, 1981.

[27] C. G. Weng and J. Poon, "A New Evaluation Measure for Imbalanced Datasets," in *Proceedings of the 7th Australasian Data Mining Conference - Volume 87*, 2008, pp. 27–32.

[28] G. Bradski, "The opencv library," *Dr. Dobb's J. Softw. Tools*, 2000.