International Journal of Innovative
Computing, Information and Control
Volume **12**, Number **5**, October **2016**

ICIC International ©2016 ISSN 1349-4198
pp. 1467–1480

VEHICLE DETECTION FOR THERMAL VISION-BASED TRAFFIC MONITORING SYSTEM USING PRINCIPAL COMPONENT ANALYSIS

Boon-Chin Yeo, Way-Soong Lim and Heng-Siong Lim

Faculty of Engineering and Technology
Multimedia University
Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia
bcyeo@mmu.edu.my

Received March 2016; revised July 2016

Abstract. *Machine vision is a popular technology used in Traffic Monitoring System (TMS) to detect vehicles in the traffic scene. Recently, thermal vision provides an alternative machine vision for the TMS since it demonstrates good vehicle detection accuracy, especially under night condition. Under the thermal vision, the vehicles appear almost similar in the daytime and nighttime, even though the illuminations of the traffic scene are significantly different. This vision effect motivates the development of a single vehicle detection algorithm that works in both of the illumination conditions. It is a challenge that has existed for decades. In this paper, a framework for thermal-vision-based TMS is proposed. Histogram of Oriented Gradients (HoG) is a feature descriptor used to recognize the vehicles on the road. The similar appearance of the vehicles under the thermal vision allows the use of single adaptive reference descriptor in vehicle detection for each lane of the road, in which the reference descriptor is generated and optimized with Principal Component Analysis (PCA). In both the daytime and nighttime, the proposed TMS framework has demonstrated high vehicle detection accuracies under the thermal vision.*
Keywords: Vehicle detection, Traffic monitoring, Thermal vision, Principal Component Analysis, Histogram of Oriented Gradients

1. **Introduction.** Vehicular traffic monitoring is the task to watch over the activities and incidents happening on the road, which can be perceived as an important infrastructure in demand for urbanization. Today, 54% of world's population lives in cities and the number is estimated to rise to 67% in 2050 [1]. Due to the urbanization process, the road networks are growing and the corresponding traffic patterns are expected to change from time to time. Instead of manually monitoring the ever-growing and massive road infrastructure over 24/7, an automated Traffic Monitoring System (TMS) is practically more convenient with minimal human intervention in its operation.

Road activities are recorded for many applications: transportation planning [2], automatic traffic control [3], traffic route planning [4], road safety analysis [5], etc. In general, the activities are recorded to study and understand the traffic demand in different regions of the road network. Vehicle count (or known as traffic volume) is a common traffic parameter generated by every TMS to basically reflect the traffic demand [6]. Considered as the derived traffic parameters from vehicle count, some TMSs are able to output the parameters such as queue length (difference between the count for vehicle entering and the count for vehicle leaving a queue) [7], and traffic flow (count for pass-by vehicle per unit time) [8]. Regardless of the traffic parameters being generated, the TMS must be able to perceive the presence of vehicles on the road in order to count them. Thus, vehicle detection can be considered as a preliminary task performed by the TMS.

Machine vision is a popular technology applied for vehicle detection in recent decades and there are large experiences in various vision-based monitoring systems whether they are indoor or outdoor. Many works have been developed for vision-based TMS to accurately detect vehicles in various environments, although the traffic scene is generally time-varying and complex [9]. Due to significantly different illumination conditions, early works pointed out the challenge to develop a single image analysis algorithm for vehicle detection that can operate both in the daytime and at night [10]. Vehicle can be detected by identifying and locating the common features appearing on every vehicle. However, the appearance of the common features is normally changing from time to time due to the varying outdoor traffic scene. In years of development, it has been generally agreed that different detection methods should be used under different illumination conditions [11]. Some of the common features that can be used to recognize a vehicle in the scene are: color [12], edge [13-15] and headlight-pair [16-18]. In recent years, image feature extraction methods such as Histogram of Oriented Gradients (HoG) [19-22], Haar-like features [23-25], Scale-invariant feature transform (SIFT) [26] and Speeded Up Robust Features (SURF) [27] have caught much attention from many researchers as the methods provide robustness in vehicle detection.

Extracted features, normally grouped in the form of single-column vector called feature vector, are assumed to contain relevant and significant information that represents an image data. Since there are many feature extraction methods, feature vector can be extended by grouping multiple extracted features as one vector or reduced by discarding the features that are less significant to represent the data [28]. Thus, a feature-based vehicle detection system is generally scalable in its feature vector formation. Machine learning has played an important role as a fast and economical method to implement the feature-based vehicle detection system to distinguish the vehicles from many dissimilar non-vehicle objects (such as trees, human, road-signs and light poles) in the traffic scene. The system can be developed and updated by respectively training and retraining a computing system whenever the datasets are available [20]. Each data in the datasets can be formed with a feature vector as the input and the label of the corresponding image (which the feature is extracted) as the output of the system.

Machine learning task that aims to infer a function mapping between the given input features and the output labels is known as supervised learning. Neural network is a popular computing system trainable in this manner for vehicle detection [28,29]. Since a vehicle detection normally operates to recognize two classes of objects (vehicle and non-vehicle), Support Vector Machine (SVM) can also be used in vehicle detection [19,24,30] SVM is a binary classifier with optimum margin separating the two classes of data that are transformed onto a hyper-plane. Generally, the larger the margin is, the more robust the classifier is. Optimization is not only important in classifying the data, but also important in selecting the features for classification. Often, the features extracted from images are dependent between one another without the knowledge of the developer. In this case, Principal Component Analysis (PCA) is an approach used for vehicle detection mainly to reduce the dimensionality of the data so that only the significant features are selected for classification [32].

Labeling a dataset is usually a manual task, which is generally tedious and error-prone. Unsupervised learning is another feasible method to implement the vehicle detection system, in which the system is designed to infer the hidden structure of the input feature vector without any label. In this way, man-made error in labeling tasks can be avoided. Besides, unsupervised learning also allows the vehicle detection system to retrain itself from time to time during its operation. This capability is essential for an outdoor monitoring system that needs to adapt itself to operate in the varying scene. Elimination of

manual labeling task can be achieved with the inclusion of auto-labeling mechanism into the system. For example, [22] uses an offline-trained SVM as a binary classifier to label the collected samples during its operations to form the new dataset from time to time. The dataset is then used to retrain the same SVM to become a more accurate classifier. Instead of retraining the offline-trained classifier, [26] uses the offline classifiers to label the new collected data to train a set of separate online classifiers during its operation.

Lately, thermal vision (which is an infrared imaging technology) has emerged as a surveillance technology for road traffic monitoring [33]. In this paper, we introduce an unsupervised-learning framework for thermal-vision-based vehicle detection using PCA. The proposed framework eliminates the need for manual data labeling in machine learning. In addition, the use of PCA allows the TMS to generate reference features, which are optimized for the vehicle detection process. The following section highlights the perspective effect of the traffic scene and vehicles under the thermal vision that motivates the approach. In Section 3, the proposed framework for traffic monitoring system is discussed in detail. Section 4 presents the results and discussions of the topic. Lastly, Section 5 concludes our findings.

2. **Perspective Effect of Thermal Vision.** Performance of vehicle detection depends not only on the choice of the algorithms, but also on the choice of the artificial sensors used for the TMS to perceive the scene environment. A conventional vision-based vehicle detection system generally relies on the use of visual camera as the sensor to perceive the outdoor scene environment. Since the camera operates in visible spectrum, many works have pointed out the problems encountered by the system to operate under varying lighting condition. During the day, shadow cast by a vehicle on the road has the tendency to be detected by the system as a part of the vehicle [34,35]. In the nighttime, low illumination of the environment, however, tends to blind the system from detecting some parts of the vehicle. In this case, the headlight-pair is often considered as an important cue for vehicle detection [11]. The upper row of Figure 1 shows the appearance of the vehicles under the normal vision in the daytime and nighttime. Vehicles in the daytime can be clearly seen, compared to the nighttime condition. In the nighttime, only the headlights are the common parts of the vehicles that can be clearly seen. The dissimilar appearances of the vehicles under different lighting conditions have become the major

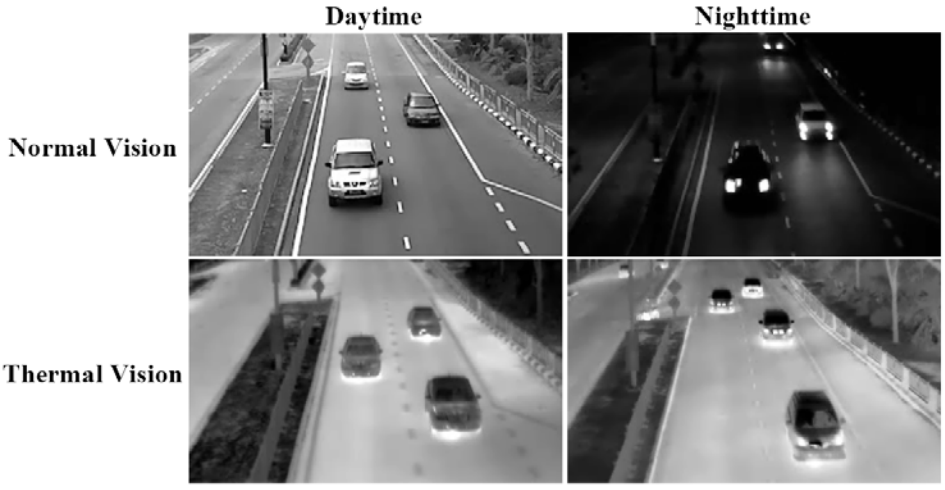

FIGURE 1. Comparison between normal vision and thermal vision in the daytime and nighttime

reason for the need to incorporate multiple detection algorithm into a vehicle detection system.

As an imaging technology that perceives thermal distribution (or infrared radiation) of objects, thermal vision does not only allow a surveillance system to 'see' in dim-lighting environment, but also in complete darkness. In the early days, line scan sensor was the choice of infrared sensor used for vehicle detection [36]. Due to the continual development of the sensor technology, thermal camera that works like a normal camera is available today to capture images and record videos of the radiations. For vehicle detection, thermal vision has been demonstrated for providing good detection accuracy under different lighting and weather conditions [32,36,37]. The lower row of Figure 1 shows the appearance of the vehicles under the thermal vision in the daytime and nighttime. It is observed that the vehicles can be apparently differentiated from the background road, regardless of the colors of the vehicles and the illumination of the scene. In addition, the appearances of the vehicles are similar in both environments. This vision effect allows a single detection algorithm to be developed to operate in both the conditions.

3. **Framework for Traffic Monitoring System.** Although the vehicles appear almost similar in different lighting conditions, there are still some differences that are consistent with the changes of the illumination. In the daytime, the thermal distribution of the engine parts of the vehicles is more evenly distributed, such that the pixel intensities around the engine parts are close to each other. In the nighttime, the front bumpers are, however, significantly brighter (or hotter) than the bonnets. Normally, the changes of the appearance under the thermal vision take hours of time since it is unlikely to have an object that changes its temperature rapidly in the traffic scene. Thus, an adaptive TMS that has the capability to learn the appearance (or technically known as the extracted features) from time to time is proposed as shown in Figure 2.

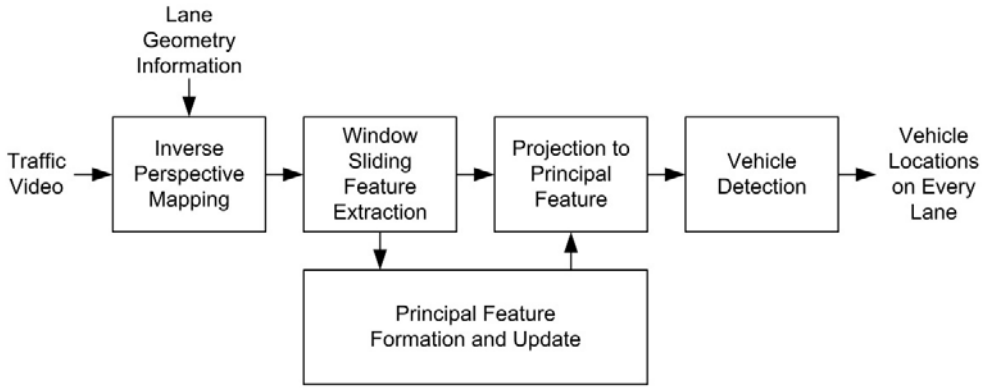

FIGURE 2. Framework for traffic monitoring system under thermal vision

Briefly, the proposed TMS analyzes the traffic video to count the vehicles on each lane. Inverse Perspective Mapping (IPM) is a preliminary process aiming to transform the perspective view of the traffic scene into the bird's-eye view. The process is essential to assure invariant vehicle's size in the captured images when the vehicle is moving on the road. The size-invariant characteristic also allows the feature extraction to be performed throughout the entire transformed image with a fixed window size. This approach generally provides benefits in terms of the computing time requirement. A reference feature will be used to detect the vehicles in the scene. From time to time, the reference feature

is updated to keep track on the changes of the appearance of the vehicles. The following subsections separately discuss the processes in detail.

3.1. **Inverse perspective mapping.** Vision-based TMS normally requires the overhead installation for the cameras. The cameras are directed to capture the traffic videos similar to the example frames presented in Figure 1. Vehicles enter from the top and leave at the bottom of the frames. Moving from top to the bottom, perspective view of the cameras causes the size of every moving vehicle to change in the traffic video. The perspective effect can be illustrated with Figure 3, in which a lane of the road laid on the xy-plane is projected onto the image represented by the uv-plane. Respectively, h and f denote the height of the camera from the ground and the focal length. In the figure, the 4 points p, q, r, and s on the xy-plane are respectively projected to p', q', r', and s' on the uv-plane. It is observed that farther lane section appears smaller than the nearer section in the image plane.

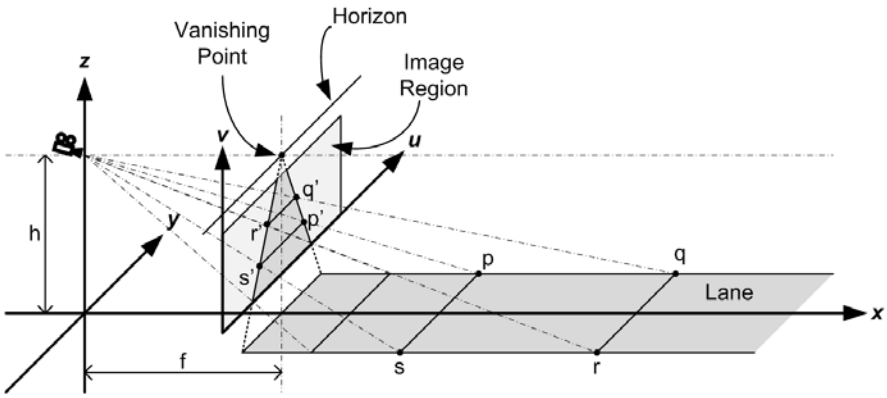

FIGURE 3. Relation between image coordinate system and the world coordinate system

To remove the scaling effect, [39] suggests that an IPM algorithm can be used to map the image from uv-plane to the xy-plane according to Equation (1). (x_i, y_i) denotes a coordinate on the xy-plane that is mapped from (u_i, v_i), which denotes a coordinate on the uv-plane. The transformation matrix A is a 3-by-3 homography matrix. Since there are 9 unknown entries, the transformation matrix A can be determined by specifying 4 sets of (x_i, y_i, u_i, v_i). In this paper, a two-lane straight road is studied. (x_i, y_i) is considered to represent the 4 vertices of a rectangle image showing the bird's-eye view of a lane. The image is hereafter called as homography image. Meanwhile, (u_i, v_i) is considered to represent the 4 vertices of a quadrilateral enclosing the target lane or Region of Interest (ROI) on the uv-plane. Figure 4 demonstrates the example homography images with different target lanes selected with quadrilaterals. It is observed that the homography images provide the bird's-eye view of the lanes. In the same homography image, the vehicles look similar to each other. However, the vehicles from different homography images consistently show slight difference in the appearances. The main reason is due to the two lanes in the scene that have different azimuths of the camera viewpoint. Vehicles in Lane 1 show more symmetry quality than the vehicles in Lane 2. The explained effect indicates that there is a need to use different features for vehicle detection in different lanes.

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = A \cdot \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}; \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1}$$

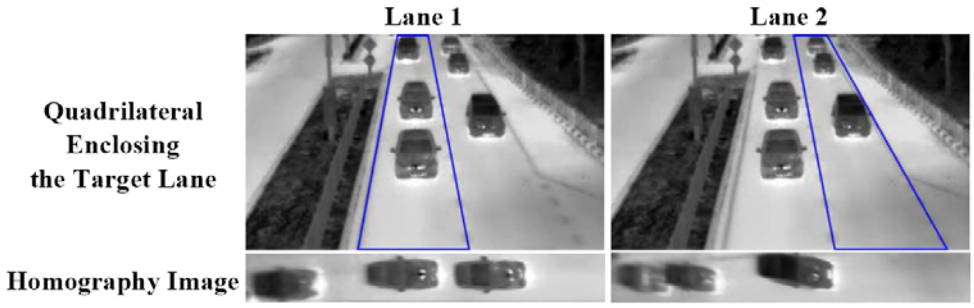

FIGURE 4. Inverse perspective mapping for different lanes

3.2. **Window sliding feature extraction.** HoG is a feature descriptor introduced by [40] for robust object detection and recognition. Recognizing objects based on the edges or gradients, HoG is thus not sensitive to the changes of the image luminance, which often results in inaccurate detection [41]. Generally, HoG descriptor is a set of numbers indicating the histogram of the local gradient orientation distribution. A homography image $I(x, y)$ can be considered as a function of a 2D surface, and the gradient $G(x, y)$ of the surface will have two perpendicular components forming a vector as shown in Equation (2). For every discrete pixel (x, y) of the homography image $I(x, y)$, the horizontal gradient $G_x(x, y)$ and vertical gradient $G_y(x, y)$ can be determined according to Equations (3) and (4) respectively. k_x and k_y are the commonly used 1D centered point derivative masks for the two directions. The magnitude gradient $|G(x, y)|$ and the gradient orientation $\theta(x, y)$ are defined with Equations (5) and (6) respectively.

$$G(x, y) = \begin{bmatrix} G_x(x, y) \\ G_y(x, y) \end{bmatrix} \tag{2}$$

where

$$G_x(x, y) = I(x, y) * k_x; \quad k_x = [-1 \quad 0 \quad 1] \tag{3}$$

$$G_y(x, y) = I(x, y) * k_y; \quad k_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{4}$$

$$|G(x, y)| = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{5}$$

$$\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \tag{6}$$

In this paper, a standard 9-bin histogram is used to form the HoG descriptor since it provides the best object detection performance as demonstrated in [40]. The 9 bins are evenly spaced over 0°-360° to include the entire regions of the gradient orientation $\theta(x, y)$. The histogram is generated for every selected image region on the homography image $I(x, y)$, using a fixed-size block or window as shown in Figure 5. The length of the window is 1.5 times the width of the homography image to allow the window to be large enough to cover the entire vehicle. Each pixel in the window casts a vote for an orientation bin according to the magnitude weighting method [42]. The method is selected because the magnitude gradient around the road region is normally significantly smaller than the magnitude gradient around the vehicle region in the homography image $I(x, y)$. The method also assures the generated descriptor of the vehicles to be apparently different from the one of the road.

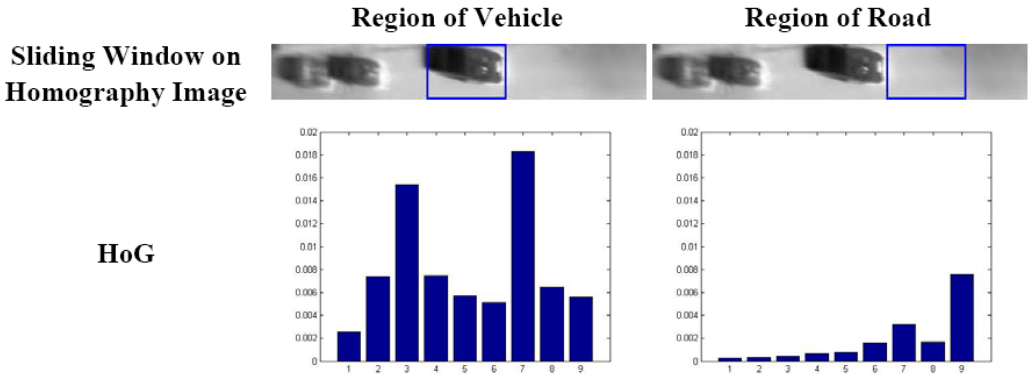

FIGURE 5. HoG of vehicle and road on homography image

3.3. **Principal feature formation, update and projection.** HoG descriptors of two similar objects will have similar values, which make the HoG suitable for object detection and recognition. Let $H(x)$ denote a vector representing an HoG descriptor as shown in Equation (7), in which $h_i(x)$ denotes the value of each bin of the histogram. Since 9-bin HoG is considered in this paper, $n = 9$. Meanwhile, x denotes the position of the center of the window sliding horizontally on the homography image. The voting mechanism to the bins can be considered to naively measure the intensity of the gradient orientation at the 9 directions individually along the x-axis. And, the use of magnitude-weight-voting presumes that the vehicles are located at the position X with high value of $h_n(x = X)$. Most importantly, $H(x = X)$ for a vehicle can be considered as a vector having different bin values as shown in Figure 5. Sliding the window away from the vehicle region will provide the vectors pointing to different directions.

$$H(x) = [h_1(x) \quad h_2(x) \quad \cdots \quad h_n(x)]^T \tag{7}$$

Figure 6 shows the density plots of the bivariate histograms generated by studying 779,015 HoG descriptors being extracted from a thermal traffic video with the frames similar to the example frames in Figure 4. For each plot, 2 bins of the descriptors are paired to vote on a 10-by-10 grid of equally spaced containers. For consistency, Bin 1 is selected for all histograms. According to the data distributions shown in the density plots, every histogram has high value near to the origin because the road background is frequently captured by the sliding window in the video. In a sufficient length of video being considered, the road region is often exposed to the camera when the vehicles are moving. Thus, the centroid (or known as the mean) of the sample descriptors can be represented as a vector B pointing to a point that is close to the origin of an n-dimensional space. Borrowing the idea from background subtraction [43], the mean vector B can be assumed as the background component representing the road region. Meanwhile, the foreground component, which shows the presence of the vehicles, can be represented with a vector $F(x)$ as shown in Equation (8).

$$F(x) = H(x) - B; \quad B = [b_1(x) \quad b_2(x) \ \cdots \ b_n(x)]^T \tag{8}$$

From Figure 6, it is also observed that there are linear dependencies between all the 9 bins. Thus, it can be assumed that the vector $H(x)$ will generally change at a unique direction when the window slides on the homography image that contains the vehicles. Let $\hat{w}$ denote the unique direction (which is also known as principal feature). The vector projection $p(x)$ of $H(x)$ to $\hat{w}$ shown in Equation (9) can be used to indicate the vehicle location. Vehicles are likely to presence at the x values with high $p(x)$. Since vector B is considered as the background component, vehicles are only expected to appear at

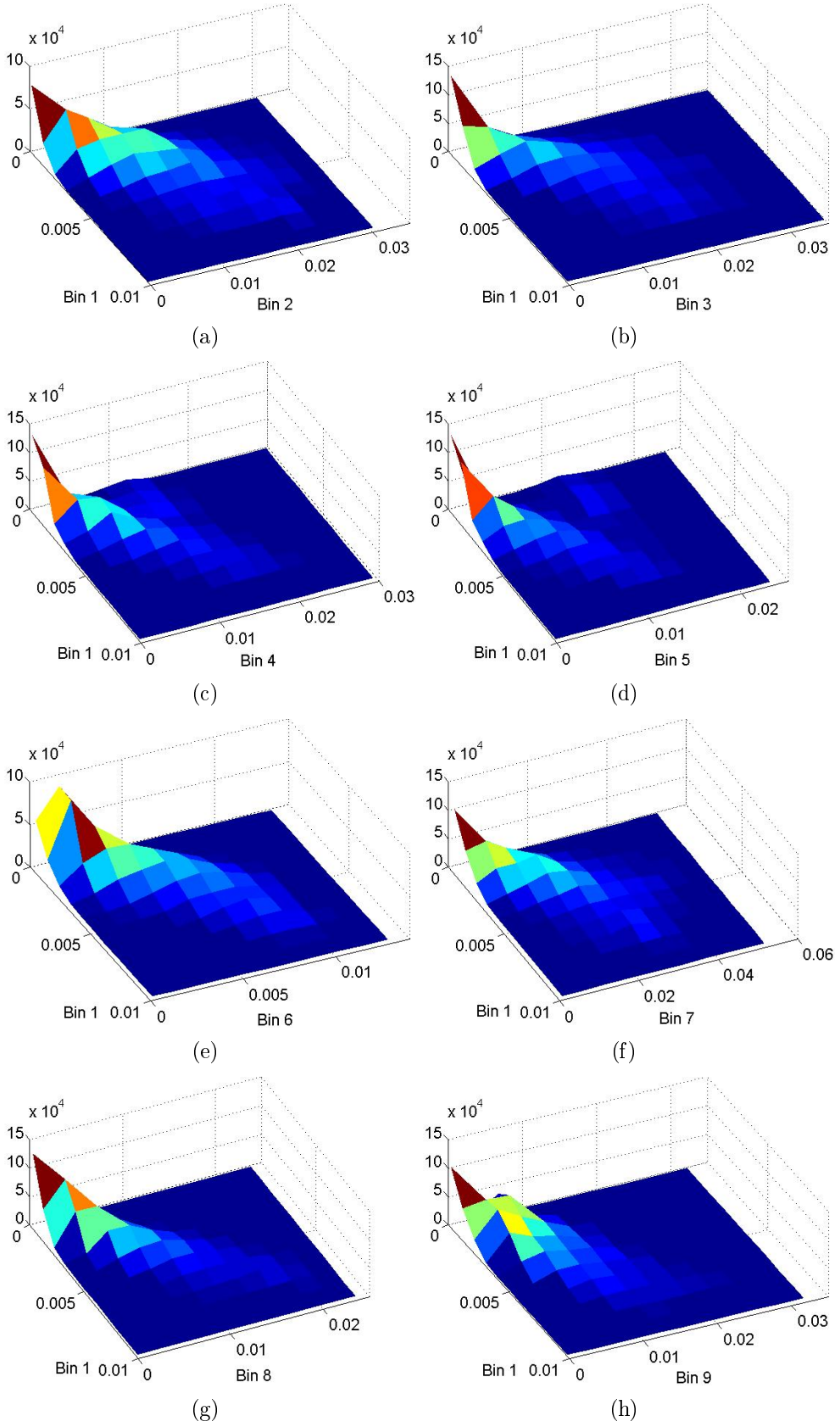

FIGURE 6. Density plot of bivariate histograms for different pairs of HoG's bins

the location with $p(x)$ greater than a threshold th. The threshold th is a projection of the vector B to the direction $\hat{w}$. To locate the vehicles with zero threshold value, an alternative vector projection $p_F(x)$ shown in Equation (10) can be used instead.

$$p(x) = \hat{w} \cdot H(x); \quad \hat{w} = [w_1 \quad w_2 \quad \cdots \quad w_n] \tag{9}$$

$$p_F(x) = \hat{w} \cdot H(x) - th = \hat{w} \cdot F(x); \quad th = \hat{w} \cdot B \tag{10}$$

The vector $\hat{w}$ can also be viewed as a direction, in which the sample descriptors show the greatest variance, compared to all other possible directions that are orthonormal to $\hat{w}$. For an n-bin HoG, there are $n-1$ vectors that are orthonormal to $\hat{w}$. Determining the direction $\hat{w}$ is a problem to be solved with PCA [44]. Consider a dataset S_D containing m number of n-bin HoG descriptors extracted from a traffic video for certain time duration. The dataset S_D can be organized into an $n \times m$ matrix. Each column of S_D is a single HoG descriptor. Meanwhile, the number of the columns of S_D is determined by m number of the descriptors being horizontally concatenated together. For example, if all the descriptors in the earlier linear dependency study are used to form the matrix S_D, the values of n and m will be 9 and 779,015 respectively. Since only the foreground component $F(x)$ of a descriptor $H(x)$ is useful to locate the vehicles, it is important to center the matrix S_D by subtracting every column of the matrix with the mean of the dataset. The mean of the dataset is also recognized as the background B.

Let S_V denote the centered dataset of S_D. PCA suggests that there is an n-by-n matrix Q transforming the dataset S_V to a new dataset S_U as shown in Equation (11). The matrix Q houses n rows of orthonormal basis vectors q_n, in which one of the basis vectors is $\hat{w}$. The matrix Q is also known as orthogonal matrix, in which $Q^{-1} = Q^T$. Defined in Equation (12) and Equation (13), let C_V and C_U denote the covariance matrices for S_V and S_U respectively. Substituting Equation (11) into Equation (13), it is found that Q^T is a matrix diagonalizing the covariance matrix C_V as shown in Equation (14), where the matrix C_U becomes a diagonal covariance matrix related to C_V. The orthonormal basis vectors q_i in the matrix Q can be now considered as the principal components or the eigenvectors. The diagonal values λ_i in the covariance matrix C_U are corresponding to the variances of the dataset S_V distributing along the direction q_i. In this context, direction $\hat{w}$ is equal to the eigenvector q_i with the corresponding variance $\lambda_i = \lambda_{\max}$ as shown in Equation (15). Since $\hat{w}$ is generated according to the samples extracted from traffic videos, it can be updated or replaced from time to time during its operation to keep track on the changes of the traffic scene.

$$S_U = Q S_V; \quad Q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \tag{11}$$

$$C_V = \frac{1}{n} S_V S_V^T = \begin{bmatrix} c_{v,11} & c_{v,12} & \cdots & c_{v,1n} \\ c_{v,21} & c_{v,22} & \cdots & c_{v,2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{v,n1} & c_{v,n2} & \cdots & c_{v,nn} \end{bmatrix} \tag{12}$$

$$C_U = \frac{1}{n} S_U S_U^T = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{13}$$

$$C_U = \frac{1}{n} Q S_V S_V^T Q^T = Q C_V Q^T \tag{14}$$

$$\hat{w} = q_i; \quad \lambda_{\max} = \max_i \{\lambda_i\} \tag{15}$$

3.4. **Vehicle detection.** The unit vector $\hat{w}$ can be considered as the reference descriptor that best represents to the vehicle on a homography image. Consequently, the vector projection $p_F(x)$ measures the similarity between the image region framed by the sliding window and a vehicle image. The higher the value of $p_F(x)$ is, the greater the similarity is. Meanwhile, lower $p_F(x)$ can be expected while sliding the window away from the vehicle region. Thus, vehicles on the homography image can be located by searching the local maximum points of $p_F(x)$. Assuming the length of a vehicle to be equal to the window length, only the local maximum points with the minimum separations equivalent to the window length are considered. To avoid getting the noisy local maximum points, $p_F(x)$ can be smoothened with a simple mean filter.

4. **Results for the Performance Tests.** In this paper, there are 2 traffic videos recorded at different hours of the day being studied. The first traffic video was captured with a thermal camera in a sunny afternoon, when the environment is hot. Figure 7 shows the

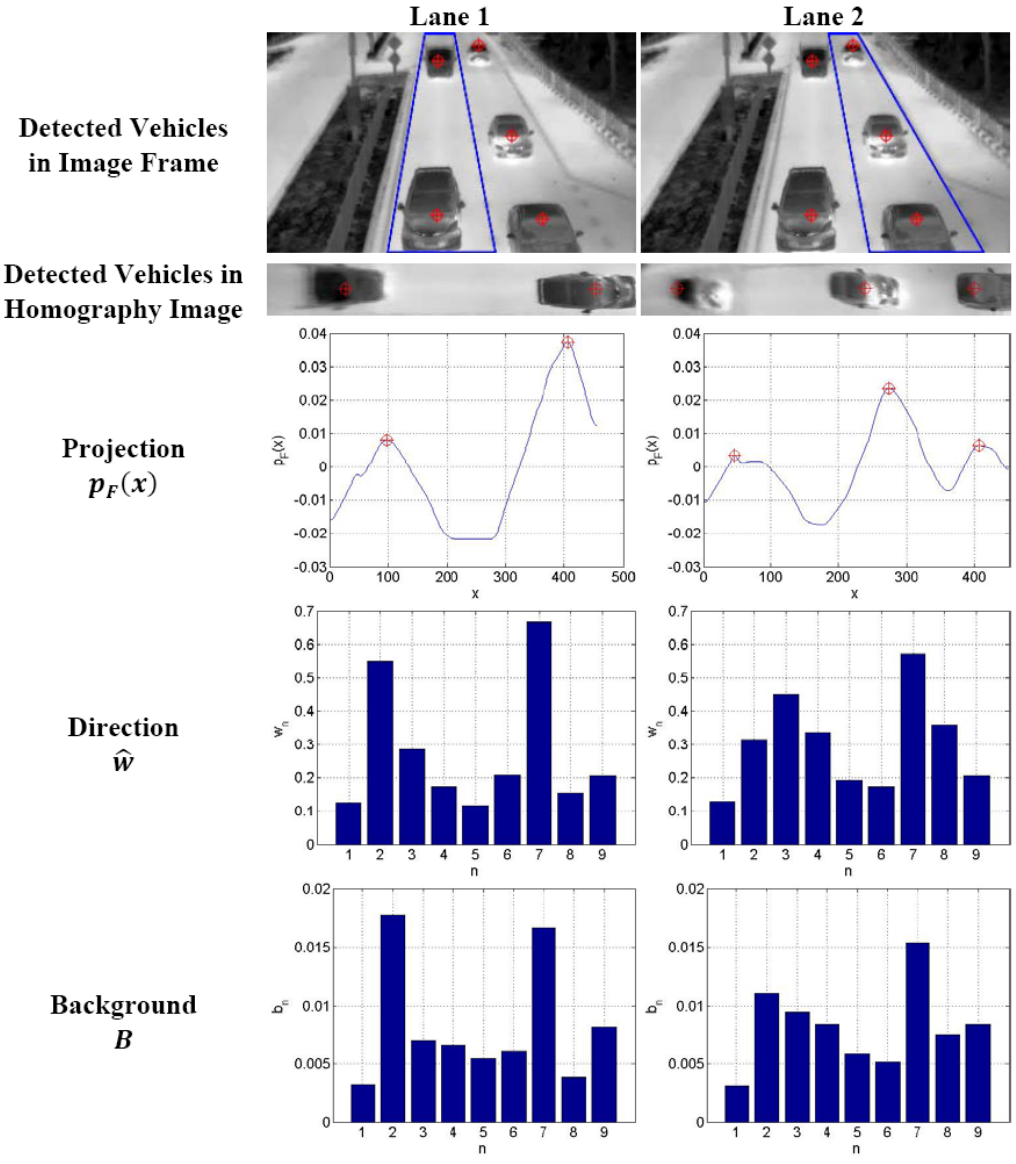

FIGURE 7. Vehicle detection in daytime

vehicles being detected in one of the video frames. The locations of the detected vehicles are marked with the marker $\oplus$ in the video frame, homography images of the two lanes, and the projection $p_F(x)$. It is observed that $p_F(x)$ is negative at the road region and a hill appears at every vehicle region. In Lane 1, the detected vehicle at the back is a truck, and the front of the truck is captured. The truck is detected due to the similar front appearance with the common vehicles like the sedan car at the front of the lane. This result demonstrates the importance of using vehicle front as the cue for vehicle detection. Even though many types of vehicles can move on the road, their front appearances are generally similar.

The direction $\hat{w}$ and background B in Figure 7 are generated according to all HoG descriptors extracted in the video. Optimized with PCA, it is observed that the vehicle detection process searches the vehicle at different direction $\hat{w}$ for different lanes. Bin 2 and Bin 7 are the two dominant bins of $\hat{w}$ for Lane 1. However, Bin 3 and Bin 7 dominate the direction $\hat{w}$ for Lane 2. For each of the lanes, the background B points to a direction similar to the direction $\hat{w}$. It can be observed by considering their dominant bins. It is mainly due to the linear dependencies between the bins as explained in Section 3.3 and the background B is a vector pointing to the centroid of the distribution.

The second video was captured with the same thermal camera in the nighttime under a dim-lighting scene at the same location. Figure 8 shows one of the video frames, where the vehicles are detected at direction $\hat{w}$. For each lane, similar directions of $\hat{w}$ are determined with PCA for vehicle detection in both the daytime and nighttime. Similar to the daytime vehicle detection, directions of $\hat{w}$ for nighttime vehicle detection are not the same for different lanes.

Accuracy is often considered as a criterion to evaluate the performance of a vehicle detection system. The accuracies for the daytime and nighttime operations are presented in Table 1. In the first videos (recorded in daytime), there are 4999 vehicles recorded. Among them, 4755 vehicles are detected, giving an accuracy of 95.12%. In the second video (recorded in nighttime), 3809 vehicles are detected among 3993 recorded vehicles, in which the accuracy is 95.39%. Thus, the PCA-optimized vehicle detection system can be used in the outdoor environment, even if the illumination of the scene changes drastically.

TABLE 1. Vehicle detection accuracy in daytime and nighttime

	Total Vehicles	**Number of Detected Vehicles**	**Accuracy**
Daytime	4999	4755	95.12%
Nighttime	3993	3809	95.39%

5. **Conclusions.** Thermal vision provides an alternative machine vision for Traffic Monitoring System (TMS) as it demonstrates good vehicle detection accuracy, especially under night condition. In this paper, an unsupervised-learning-based vehicle detection algorithm for traffic monitoring system under thermal vision is developed. Histogram of Oriented Gradients (HoG) is a feature descriptor used to recognize the vehicles on the road. The appearances of vehicles in the same lane are almost similar to each other under the thermal vision, regardless of the colors of the vehicles and the illumination of the traffic scene. The thermal vision effect results in similar HoG extracted for all the vehicles. Also, the HoG extracted along the lane shows the existence of linear dependencies between all the bins of the HoG descriptor. In this context, a single adaptive reference descriptor can be used for vehicle detection for each lane, in which the reference descriptor is generated and

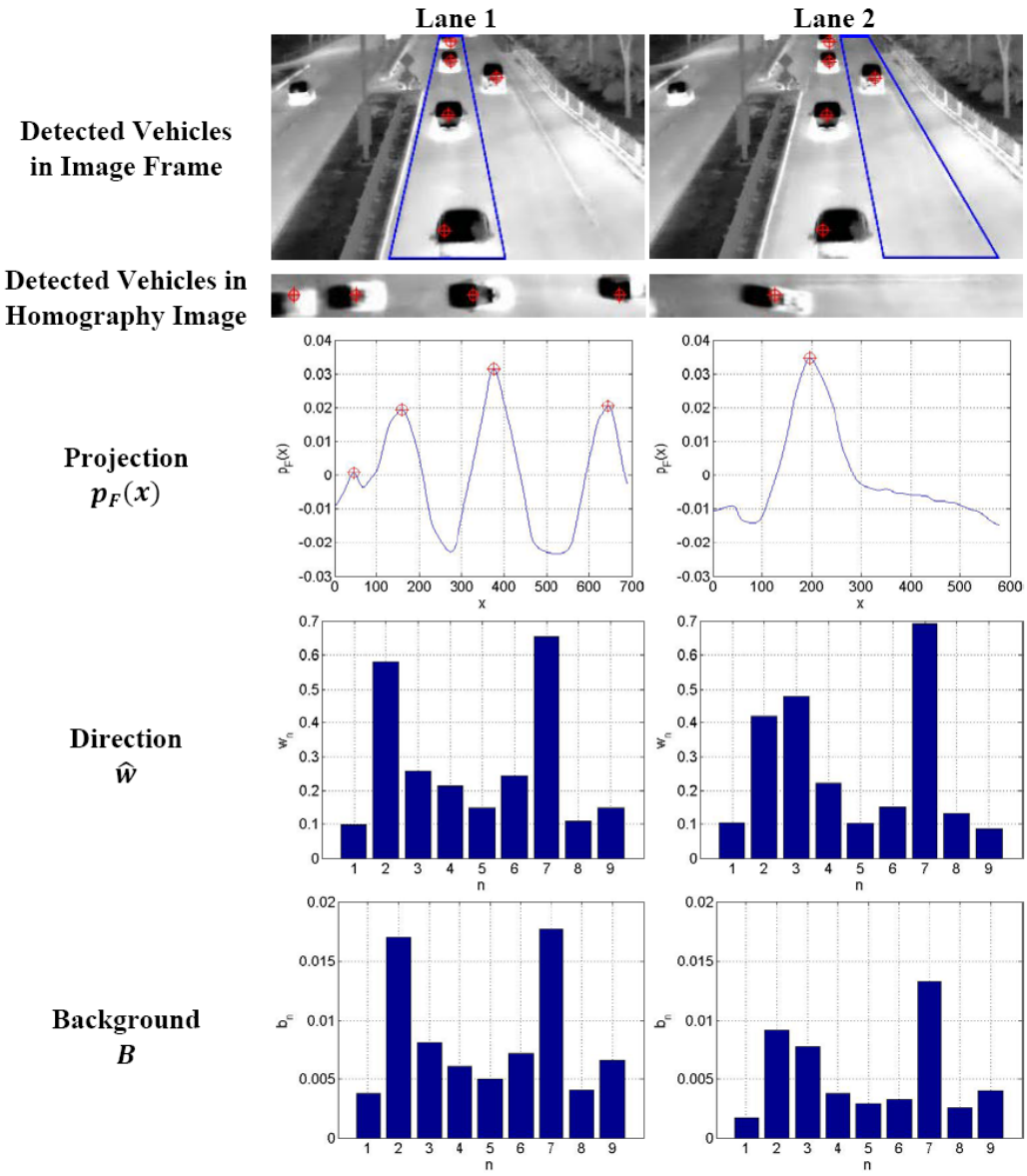

FIGURE 8. Vehicle detection in nighttime

optimized with Principal Component Analysis (PCA). In both the daytime and nighttime, when the illuminations of the traffic scene are significantly different, the proposed TMS framework has demonstrated more than 95% accuracy for vehicle detection under the thermal vision.

At the moment, only the daytime and nighttime environments are being studied with the proposed framework. Future works may focus on studying the proposed framework for more weather conditions like raining and fog. In addition, the IPM module of the framework may also be revised in the future with a different camera setup to enable wider viewing scope for the TMS and make the TMS more economical.

REFERENCES

[1] United Nations, *World Urbanization Prospects*, The 2014 Revision, New York, 2015.
[2] K. Sigakova, G. Mbiydzenyuy and J. Holmgren, Impacts of traffic conditions on the performance of road freight transport, *IEEE International Conference on Intelligent Transportation Systems*, pp.2947-2952, 2015.

[3] S. Tahilyani, M. Darbari and P. K. Shukla, Soft computing approaches in traffic control systems: A review, *AASRI Procedia*, vol.4, pp.206-211, 2013.
[4] J. Aslam, S. Lim and D. Rus, Congestion-aware traffic routing system using sensor data, *International IEEE Conference on Intelligent Transportation Systems*, pp.1006-1013, 2012.
[5] Z. Christoforou, S. Cohen and M. G. Karlaftis, Integrating real-time traffic data in road safety analysis, *Procedia – Soc. Behav. Sci.*, vol.48, pp.2454-2463, 2012.
[6] A. S. Othman, Y. Burette and H. H. Refai, Low cost upgrade for established traffic monitoring systems to support real-time, *The 12th Annual IEEE Consumer Communications and Networking Conference*, pp.886-891, 2015.
[7] A. Wu and X. Yang, Real-time queue length estimation of signalized intersections based on RFID data, *Procedia – Soc. Behav. Sci.*, vol.96, pp.1477-1484, 2013.
[8] L. Xiao, P. Peng, X. Wang, Z. Xu and B. Hong, Research on traffic monitoring network and its traffic flow forecast and congestion control model based on wireless sensor networks, *International Conference on Measuring Technology and Mechatronics Automation*, pp.142-147, 2009.
[9] S. Sivaraman and M. M. Trivedi, A review of recent developments in vision-based vehicle detection, *IEEE Intelligent Vehicles Symposium*, pp.310-315, 2013.
[10] R. Cucchiara and M. Piccardi, Vehicle detection under day and night illumination, *Proc. of ISCS-IIA99*, 1999.
[11] J. Guo, J. Wang, X. Guo, C. Yu and X. Sun, Preceding vehicle detection and tracking adaptive to illumination variation in night traffic scenes based on relevance analysis, *Sensors*, vol.14, no.8, pp.15325-15347, 2014.
[12] L. Tsai, J. Hsieh and K. Fan, Vehicle detection using normalized color and edge map, *IEEE International Conference on Image Processing*, vol.2, pp.598-601, 2007.
[13] B. F. Lin, Y. M. Lin, L. C. Fu, P. Y. Hsiao, L. A. Chuang, S. S. Huang and M. F. Lo, Integrating appearance and edge features for sedan vehicle detection in the blind-spot area, *IEEE Trans. Intell. Transp. Syst.*, vol.13, no.2, pp.737-747, 2012.
[14] W. Tsai, S. Wu, L. Lin, T. Chen and M. Li, Edge-based forward vehicle detection method for complex scenes, *IEEE International Conference on Consumer Electronics – Taiwan*, pp.173-174, 2014.
[15] G. Y. Song, K. Y. Lee and J. W. Lee, Vehicle detection by edge-based candidate generation and appearance-based classification, *IEEE Intelligent Vehicles Symposium*, pp.428-433, 2008.
[16] Q. Zou, H. Ling, S. Luo, Y. Huang and M. Tian, Robust nighttime vehicle detection by tracking and grouping headlights, *IEEE Trans. Intell. Transp. Syst.*, vol.16, no.5, pp.2838-2849, 2015.
[17] H.-Y. Cheng, P.-Y. Liu and Y.-J. Lai, Lane departure system design using with IR camera for night-time road conditions, *International Conference on Education Technology and Computer*, vol.5, pp.122-125, 2010.
[18] W. Zhang, Q. M. J. Wu and G. Wang, Vehicle headlights detection using Markov random fields, *Computer Vision – ACCV 2009*, 2009.
[19] J. Kim, S. Hong, J. Baek, E. Kim and H. Lee, Autonomous vehicle detection system using visible and infrared camera, *The 12th International Conference on Control, Automation and Systems*, pp.630-634, 2012.
[20] H. Wang and H. Zhang, A hybrid method of vehicle detection based on computer vision for Intelligent Transportation System, *Int. J. Multimed. Ubiquitous Eng.*, vol.9, no.6, pp.105-117, 2014.
[21] X. Yu and Z. Shi, Vehicle detection in remote sensing imagery based on salien information and local shape feature, *Optik (Stuttg)*, vol.126, no.20, pp.2485-2490, 2015.
[22] B. Tamersoy and J. K. Aggarwal, Robust vehicle detection for tracking in highway surveillance videos using unsupervised learning, *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp.529-534, 2009.
[23] X. Bao, S. Javanbakhti, S. Zinger, R. Wijnhoven and P. H. N. de With, Context-based object-of-interest detection for a generic traffic surveillance analysis system, *The 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pp.136-141, 2014.
[24] X. Wen, L. Shao, W. Fang and Y. Xue, Efficient feature selection and classification for vehicle detection, *IEEE Trans. Circuits Syst. Video Technol.*, vol.25, no.3, pp.508-517, 2015.
[25] R. Rios-Cabrera, T. Tuytelaars and L. van Gool, Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application, *Comput. Vis. Image Underst.*, vol.116, no.6, pp.742-753, 2012.
[26] Z. Qian, H. Shi, J. Yang and L. Duan, Video-based Multiclass vehicle detection and tracking, *Int. J. Comput. Sci. Issues*, vol.10, no.1, pp.570-578, 2013.

[27] J. W. Hsieh, L. C. Chen and D. Y. Chen, Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition, *IEEE Trans. Intell. Transp. Syst.*, vol.15, no.1, pp.6-20, 2014.
[28] J. Arrospide and L. Salgado, A study of feature combination for vehicle detection based on image processing, *Sci. World J.*, vol.2014, pp.1-13, 2014.
[29] P. M. Daigavane, P. R. Bajaj and M. B. Daigavane, Vehicle detection and neural network application for vehicle classification, *International Conference on Computational Intelligence and Communication Networks*, pp.758-762, 2011.
[30] X. Song, T. Rui, Z. Zha, X. Wang and H. Fang, The Adaboost algorithm for vehicle detection based on CNN features, *Proc. of the 7th International Conference on Internet Multimedia Computing and Service*, 2015.
[31] Z. Sun, G. Bebis and R. Miller, On-road vehicle detection using Gabor filters and support vector machines, *International Conference on Digital Signal Processing*, vol.2, pp.1019-1022, 2002.
[32] J. Arrospide and L. Salgado, Region-dependent vehicle classification using PCA features, *IEEE International Conference on Image Processing*, pp.453-456, 2012.
[33] Y. Iwasaki, M. Misumi and T. Nakamiya, Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring, *Sensors*, vol.13, no.6, pp.7756-7773, 2013.
[34] Z. Kim and J. Malik, Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking, *The 9th IEEE International Conference on Computer Vision*, pp.524-531, 2003.
[35] Y. Li, Z. Li, H. Tian and Y. Wang, Vehicle detecting and shadow removing based on edged mixture Gaussian model, *IFAC World Congress*, pp.800-805, 2011.
[36] P. G. Ducksbury, D. M. Booth and C. J. Radford, Vehicle detection in infrared linescan imagery using belief networks, *The 5th International Conference on Image Processing and Its Applications*, pp.415-419, 1995.
[37] C. Antonellot, A. Broggil, A. Fascioli and D. Ranzato, Vehicle detection and localization in infra-red images, *The 5th International Conference on Intelligent Transportation Systems*, pp.141-146, 2002.
[38] A. Sangnoree and K. Chamnongthai, Robust method for analyzing the various speeds of multitudinous vehicles in nighttime traffic based on thermal images, *The 4th International Conference on Computer Sciences and Convergence Information Technology*, pp.467-472, 2009.
[39] S. Tuohy, D. O'Cualain, E. Jones and M. Glavin, Distance determination for an automobile environment using inverse perspective mapping in OpenCV, *IET Irish Signals and Systems Conference*, pp.100-105, 2010.
[40] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp.886-893, 2005.
[41] X. Cao, C. Wu, P. Yan and X. Li, Linear SVM classification using boosting hog features for vehicle detection in low-altitude airborne videos, *The 18th IEEE International Conference on Image Processing*, pp.2421-2424, 2011.
[42] J. Nilsson and H. Peter, Gradient sample argument weighting for robust image region description, *IEEE International Conference on Electronics, Computing and Communication Technologies*, pp.1-4, 2013.
[43] B. C. Yeo, W. S. Lim and H. S. Lim, Scalable-width temporal edge detection for recursive background recovery in adaptive background modeling, *Appl. Soft Comput.*, vol.13, no.4, pp.1583-1591, 2013.
[44] J. Shlens, *A Tutorial on Principal Component Analysis*, http://arxiv.org/abs/1404.1100, 2014.