International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 2
ISSN: 2088-5334

A Semantic Extraction and Analysis for Traffic Density Using Traffic Images: A Critical Review

Ruhana Abang Yusup[a,*], Wang Hui Hui[a], Wee Bui Lin[a]

[a] *Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia*
Corresponding author: [*]ruhanaay@gmail.com

***Abstract*—Population growth in large cities has contributed to the increase in vehicles' number, leading to the traffic congestion problem. Incompetent traffic supervision could squander an inconsiderable number of man-hours and might lead to fatal consequences. Therefore, intelligent traffic surveillance systems have to carry more significant roles in highway monitoring and traffic management system throughout the years. Although vehicle detection and classification methods have evolved rapidly throughout the years, they still lack high-level reasoning. Accurate and precise vehicle recognition and classification are still insufficient to develop an intelligent and reliable traffic system. There is a demand to increase the confidence in image understanding and effectively extract the images conformed to human perception and without human interference. This paper attempts to summarize a review on several methods that semantically extract and analyze traffic density with image processing techniques. Three (3) methods that have been selected to be discussed in this paper are semantic analysis of traffic video using image understanding, mining semantic context details of traffic scene, and integrating vision and language in semantic description of traffic events from image sequences. Each method is discussed thoroughly, and their outstanding issue is deliberated in this paper.**

***Keywords*— Intelligent traffic surveillance; semantical analysis; traffic images; traffic density.**

Manuscript received 24 Oct. 2019; revised 25 Dec. 2020; accepted 3 Mar. 2021. Date of publication 30 Apr. 2021.

I. INTRODUCTION

The rapid development in the social economy today has resulted in the improvement of lifestyle and living levels. Most of the population could afford various means of transportation. Therefore, traffic congestion has been increasing lately in many development areas due to the rise in vehicles. Although other research has been done over the years to develop the most intelligent traffic surveillance system, the combination of some approaches such as content-based image and video retrieval technology is still unable to express exact and complete high-level semantic results. In other words, they still lack high-level reasoning that is required to understand the significance of the traffic objects or scenes and the meaning its conveyed.

Nonetheless, several research types have been carried out, which attempted to analyze or describe traffic images semantically. Some approaches intended to employ the image understanding method and mining the semantic context information to analyze traffic video semantically. Meanwhile, other procedures discuss the implementation of vision and language integration to provide traffic image descriptions semantically. The use of ontologies to cover traffic occurrences using vocabularies to annotate traffic video resources is also included in some studies. These approaches will be reviewed and compared throughout this paper to come up with the most practical approach.

The traffic surveillance system has evolved over the years. This evolution can be categorized into three simple stages. The first stage is the employment of several approaches that do not implement image processing techniques, such as Magnetic Loop Detector (MLD) and infra-red sensor. Magnetic Loop Detector (MLD), with the help of its magnetic traits, is buried under the road to calculate the vehicle's quantity, whereas the infra-red sensor is installed at the side of the road to monitor traffic flow [1]. However, Magnetic Loop Detector (MLD) provides limited traffic information and requires a discrete vehicle counting and traffic surveillance system [1]. On the other hand, the infra-red sensor is subjected to a high failure rate in the situation where fogs and mists are present [2]. Other than that, the traffic surveillance system also employed an inductive loop detector that seems to be a costlier solution but unable to deliver accurate and consistent results when installed in a deprived

road surface condition [1]. Not to mention, it might also interrupt traffic flow throughout restoration and maintenance [3]. Besides that, light beams such as LASER and IR are also adopted as traffic surveillance system approaches. However, as traffic moves, light beams are obstructed [3]. The authors also proposed an acoustic sensor that estimates road congestion by analyzing road noise from a vehicle-mounted microphone attached outside the vehicle [29].

Nevertheless, acoustic sensors are only functional in a short-range distance and prone to interference problems due to noisy traffic environments. A radar-based vehicle detection system has been proposed where input data are obtained from radar sensors to be processed using signal processing unit to distinguish the vehicle [35]. Although radar sensors can be considered a robust detection system because they are not affected by environmental challenges, they can still provide detailed information about the vehicle, such as shape, size, texture, and color [26]. Others attempted to combine the sensors to maximize their functionality; however, it required separate algorithms for each sensor and is quite expensive to be implemented [29].

Due to the issue that comes with the previous approaches in the first stage, a more contemporary approach that implemented image processing technique is recommended in the second stage. The image processing technique is a cost-effective approach involving processing digital images employing a computer to provide broad information and more reliable data. These progressions in computer vision techniques and machine learning algorithms pave the way for creating novel algorithms capable of detecting and counting the vehicles and classifying the type of vehicles [11]. In addition, these approaches can easily support information feed through telephone or web networks. Cameras for observing traffic scene will be installed on each side of the traffic junction [4]. Next, traffic parameters and information that the video camera has captured will be broadcast to the servers, where it will be processed using image processing techniques to extract the real-time traffic information [5]. There are several object detection and recognition techniques that traffic images have to encounter to extract the traffic information. It has been suggested that traffic images go through background subtraction and canny edge detection before counting the number of vehicles [4]. Background subtraction is a conventional method that computes the error between constant background frames with the current one for real-time segmentation of an object in a video-based system [4], [11], [25]. As for canny edge detection, it is an approach where the images are smoothed to find the image gradient that will highlight regions with high spatial derivatives [4]. The advantages of adopting the canny edge detection method are reducing the signal-to-noise ratio, reducing multiple responses into one and only edge, and guaranteeing the edge points are localized thoroughly [33]. Besides, it has been proposed that the color features extraction method is integrated with a line detection technique for object identification and representation in complex traffic scenes [6]. However, another study comes up with a novel approach for vehicle detection where they extract and analyze gradient and range features on detection lines that they can obtain by installing virtual line-based sensors on highway lanes [21]. Like the previous method, Tang et al [13] also employed features extracting method via Haar-like features and adopt AdaBoost algorithms to construct classifiers to determine the position of the vehicle in the image. Apart from that, Tang *et al.* [13] also use Gabor wavelet transform and local binary operator to extract multi-scale and multi-orientation vehicle features before applying principal component analysis (PCA) and Euclidean distance comparison algorithm for vehicle type recognition. Gabor wavelet transform is a reliable technique compared to others, especially in identifying the vehicle's size due to its multi-resolution and multi-orientation properties and its robustness against noise various illumination [31]. As for Wen *et al.* [32] studied, the authors also suggested the method based on Haar-like features with AdaBoost algorithm but with a slight change introducing an improved normalization algorithm for vehicle detection. Apart from that, Abid *et al.* [19] presented an image detection technique based on multi-scale covariance (MSCOV) descriptor for image description and support vector machine (SVM) classifier for vehicle classification. Support vector machine (SVM) is a superior classifier in classifier-based vehicle verification method where two-group image classifiers are implemented to distinguish the vehicle from non-vehicle [29]. Wang *et al.* [28], on the other hand, proposed the Improved Spatio-Temporal Sample Consensus technique to improve background detection technique in order to distinguish moving vehicles despite the intrusion of their shadow and light variation and categorized them by using a multi-feature fusion approach. Liu and Mattyus [30] suggested the use of binary sliding window detector based on integral channel features (ICFs) and AdaBoost classifier to generate the vehicle bounding box before applying a histogram of oriented gradients (HOG) features to identify the type and orientation of the vehicle in aerial traffic images. Another paper proposed a non-linear technique using the multi-scale differential morphological profile for vehicle detection where the vehicle's shape index is utilized to identify the type of vehicle in traffic images [14]. While some studies [11], [27], [31] implemented image processing techniques for vehicle counting, which are background subtraction technique, texture analysis via Gray-Level Co-Occurrence Matrix (GLCM) calculation, and feature extraction via Gabor wavelet transform, respectively. However, traffic density estimation is successfully done by removing the vehicle detection step completely and replacing the method with block-based processing approach where each road lane is partitioned into multiple blocks and the percentage occupancy of the lanes is computed by identifying which block are occupied by the vehicle [15].

The difficulty of object detection and classification escalates with different variables such as orientation, size, and figure, not to mention that obstacles such as illumination variation, noise, shadow, and inaccuracy during segmentation might affect the result [16]. Therefore, some authors proposed a more reliable technique in deep learning to accurately identify and classify vehicles. Deep learning is an approach that studies distinct features simply from input images for a particular mission in a controlled manner [36]. Haeikki *et al.* [12] proposed two data-driven frameworks which consist of deep neural networks and support vector machines (SVM) via Scale Invariant Feature Transform (SIFT) for vehicle type recognition. Even though the result shows 97% accuracy in

detecting four bus, truck, van, and car classes, classification errors still occur when the vehicles are considered in the middle of two classes, such as minivan [12]. A paper implemented a semi-supervised convolutional neural network for vehicle type recognition where the SoftMax classifier function as the output layer, and its filters are studied using the Laplacian filter learning approach [17]. Not to mention, others suggested the collaboration between convolutional neural network (CNN) and recurrent neural network (RNN) to produce accurate description for traffic images based upon attention mechanism [18]. Other than that, Audebert *et al.* [20] proposed a segment before detecting approach where the datasets are trained using a deep, fully convolutional network in order to study the semantic maps to obtain the accurate and exact segmentation of vehicles and then categorize the vehicles with the help of convolutional neural network (CNN). As for Suhao *et al.* [22], the authors suggested the Faster Recurrent Neural Network (Faster-RCNN) framework by inserting the sample images into an improved Region Proposal Network (RPN) training before entering the convolutional network parameters (trained by RPN) in the Fast-RCNN network to identify three types of vehicle which are car, minibus, and SUV. Apart from that, Zhou and Cheung [23] proposed the employment of lane marking to identify the vehicle's position in the rear-view vehicle images before applying Deep Neural Network (DNN) for vehicle classification. Vijayaraghavan and Laavanya [24], on the other hand, produce their own novel version of convolutional neural network (CNN) capable of detecting three types of vehicle such as bus, car, and motorcycle with contemporary CNN formulated on fast regions.

The authors believed that if the convolutional neural network is proficient enough to mine the features precisely, the region-based detector will identify the object. Besides, Tsai *et al.* [38] also employed an improved CNN by fine-tuning its current framework. The authors implement hypernet architecture with eight inception layers for the base network and eight Concatenated ReLU convolution layers to precisely generate the vehicle bounding box. Last but not least, Arinaldi *et al.* [37] employed two models for vehicle recognition and classification, which are background subtraction techniques based on a mixture of Gaussian (MoG) with support vector machine (SVM) and faster region-based convolutional neural network (Faster R-CNN) in order to compare their performance. The end result proves that faster R-CNN surpasses MoG with SVM performance in identifying the vehicle during low illumination conditions and handling multiple vehicle occlusion situations.

Even though various research have been carried out to come up with the most effective image processing technique to analyze traffic scene images, they still lack high-level semantic reasoning. Traffic images that are stored in the majority of image retrieval systems were described by using low-level features. If high-level semantic offer complete and accurate image content description, low-level portray the opposites. Thus, extraction of traffic image in abstract attributes get little attention even though they may have crucial information on objects or scenes due to insufficient high-level reasoning. Therefore, in the third stage, researchers strove to develop various approaches that propose semantic extraction and representation techniques to describe traffic events and derived traffic information. The next part of this paper attempted to summarize and review several of those existing approaches.

II. Materials and Method

A. Semantic Analysis of Traffic Video using Image Understanding

This research suggested hierarchical structure as the underlying fundamental for image understanding techniques to be implemented in the traffic video semantic analysis. Image understanding is a technique that takes knowledge from the core, studies the content of the image and its relationship, comprehends its scene, and decides how to use the scenarios [7]. Since the vehicle status analysis outcomes are the essential component in traffic incident detection because it can directly affect the application level of traffic video analysis, this approach recommended video semantic analysis framework through image understanding due to its instantaneous (real-time) and precise evaluation [7].

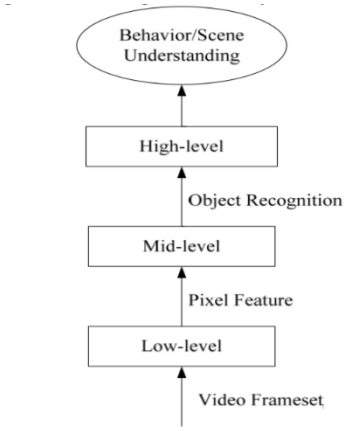

Fig. 1 Video image understanding mechanism [7]

There is a hierarchy to comprehend the course of video image understanding, and it is shown in Fig. 1. The lowest level of the hierarchy consists of video image segmentation that is in charge of separating the foreground from a background video sequence. The interlayer comprises object identification and allocation that is accountable for detecting and classifying an object in the foreground areas, whereas the uppermost layer contains behavior or scene analysis that studies the object behavior and describes the scene's state [7].

Fundamentally, the first step in this approach is utilizing the background subtraction technique to distinguish vehicles in traffic scenes. Background subtraction technique, alias foreground detection algorithm, is usually selected for an image that is a fragment of video stream due to its ability to identify moving objects from the contrast amid the current frame and reference frame, also called frame differencing [11], [25]. Next, a fast normalized cross-relation method that depends on assumption will be used to obtain the precise vehicle trajectory from the identified vehicles. After that, the vehicle's particulars such as type, color, and I/O time will be mined. Finally, with the help of the structured data above, valuable information can be extracted through statistical analysis and video querying.

1) Vehicle Tracking: The proposed approach employed a fast normalized cross-correlation method that relies on assumption to track moving objects and obtain its precise trajectory [7]. Wu *et al.* [7] fast normalized cross-relation method is a statistical estimation method that is usually applied in template matching and pattern recognition. A tracked object's moving trajectory can be acquired in the vehicle tracking process to estimate where the object in motion is positioned in the consecutive frame. Therefore, moving templates and estimated regions should be normalized as soon as possible to minimize the duration of the fast-normalized cross-correlation process [7].

2) Video Content Extraction: The first approach is Vehicle Type Recognition. Through this approach, vehicle type can be identified based on driveline detection that will categorize the vehicles into four groups which are truck, bus, van, and car. Wu *et al.* [7] mentioned that the first step to extract vehicle type is to obtain a background image from the current video frame. Thus, the background subtraction method will be employed to acquire the moving regions. However, prior to conducting the procedure, the background model should be constructed [7]. It is safe to assume that only slight greyscale alteration in the background for some moments, even though the prospect's greyscale differs a lot to numerous vehicles. After that, the second step that needs to be done is detecting two drivelines using Hough Transform [7]. As Takeuchi *et al*. [34] mentioned, Hough transform can identify patterns effectively even in the presence of noise or occlusions. Wu *et al.* [7] found that the Hough Transform simplifies line detecting process. Hough Transform is able to detect ellipses, circles and other general graphics by mapping a set of points of some graphic to one point [34]. The next step is to measure the vehicle's actual region using the ratio of measured driveline width and actual width. Lastly, the vehicle is categorized based on the actual region value of the vehicle.

Second approach is Color Recognition. To identify each vehicle's color, color information should be mined and grouped into categories [7]. The first step that needs to be done is choosing the appropriate color space. Next, the color characteristic should be defined and computed. After that, the resemblance of the color features should be extracted and matched as soon as possible. Even though RGB color space is employed widely in the images, there is an issue to use Euclidean distance to describe the distance between two different colors. Furthermore, since the RGB value will have non-linear change whenever the color changes, it is not appropriate for color classification and computation. Therefore, in this approach, Wu *et al.* [7] suggested that RGB color space should be transformed into HSL color space. Based on the L component's value, this method will differentiate the color black, white, and grey and then define the other colors according to the H component's value. As a result, HSL color can be grouped into seven (7) colors: white, grey, black, green, blue, yellow, and red, which covers most of the vehicle's base color practically.

3) Experimental Analysis

Vehicle Tracking: The proposed method is tested using a traffic video that was shot in the tunnel of DuShuHu in Suzhou. Based on Fig. 2, the fifth vehicle entered the tracking range at the 624th frame and exited at 657th frame. Correspondingly, as shown in Fig. 3, the ninth vehicle shows a van entered the tracking range at 931st frame and exited at 960th. Even though the vehicles' form and magnitude change once the vehicles move closer to the camera and leave the tracking range, the method can still detect the vehicles accurately in actual time.

Fig. 2 Two different frame when the 5th vehicle entered tracking range [7]

Fig. 3 Two different frames when the 9th vehicle entered tracking range [7]

Video Content Extraction: After a successful tracking process, vehicle information is derived. The authors categorized the information into vehicle number, entered frame number, exited frame number, vehicle type, and color. As shown in Fig. 4, the method successfully identifies the preferred vehicles' data and records the database's information.

cameraID	videoID	vehicleID	frameIn	frameOut	vehicleType	vehicleColor	memo1
0101	2009072806	1	167	184	Van	White	
0101	2009072806	2	661	703	Car	Black	
0101	2009072806	3	719	726	Car	White	
0101	2009072806	4	924	957	Car	Black	
0101	2009072806	5	1042	1042	Car	Black	
0101	2009072806	6	1066	1066	Car	Sliver	
0101	2009072806	7	1231	1260	Van	Blue	
0101	2009072806	8	1745	1772	Van	Sliver	
0101	2009072806	9	2094	2103	Car	Blue	
0101	2009072806	10	2339	2377	Car	Blue	
0101	2009072806	11	2419	2448	Car	Black	
0101	2009072806	12	2614	2657	Car	Black	
0101	2009072806	13	2623	2623	Car	White	
0101	2009072806	14	2673	2708	Car	Black	
0101	2009072806	15	2940	2940	Car	White	

Fig. 4 Extraction result of video content [7]

Video Querying and Statistical Analysis: The preferred data of the vehicle that has been recorded is converted into text to make it possible for the user to search for specific information for their specific needs [7]. Users must set the searching criteria based on date, time, vehicle type, or vehicle color. As a result, the system will provide the frameset and its corresponding video frames instantaneously. In addition, the user could also access the statistic and percent of vehicle category and vehicle color through the system just by setting specific dates and times. Wu *et al.* [7] pointed out that the

vehicle category statistic can determine the current road circumstances, whereas vehicle color statistic can indicate the public preference of vehicle color to assist the manufacture and trade company in making a strategic decision. Fig. 5(a) demonstrates the vehicle types' ratio diagram, whereas Fig. 5(b) illustrates ratio diagram of vehicle colors on 28th July 2009.

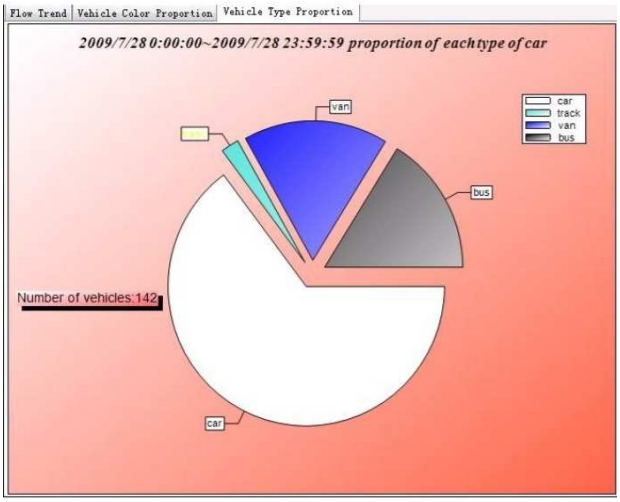

Fig. 5(a) Ratio diagram of vehicle types [7]

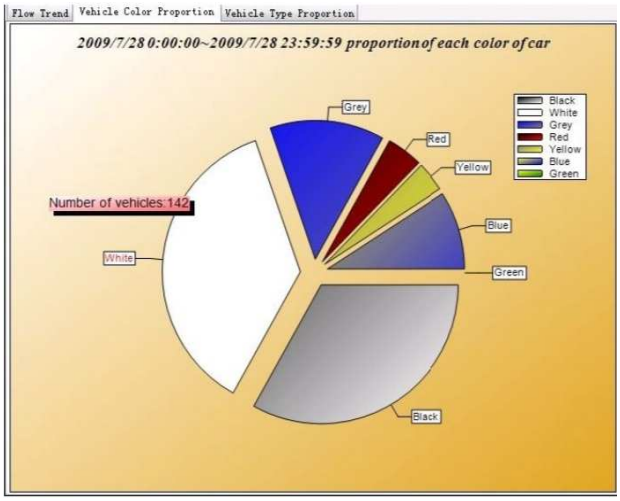

Fig. 5(b) Ratio diagram of vehicle colors [7]

B. Mining Semantic Context Details of Traffic Scenes

This proposed method attempts to enhance abnormal event detection, object detection, object classification, and object tracking in intelligent traffic surveillance systems through semantic context details derivation [8]. As stated by *Zhang* et al. [8], scene-specific context details will be reaped and studied from object-specific context details before integrating both of this information as semantic context details. Object-specific context details include speed, motion direction, aspect ratio, occupancy rate, the region in pixels, and x-, y-image coordinates [8]. Meanwhile, scene-specific context details can be attained through understanding motion pattern and width distribution from the Gaussian mixture model, studying paths through graph cut, and examining sources or sinks via a mean-shift approach [8]. Fig. 6 shows the architecture of the proposed method.

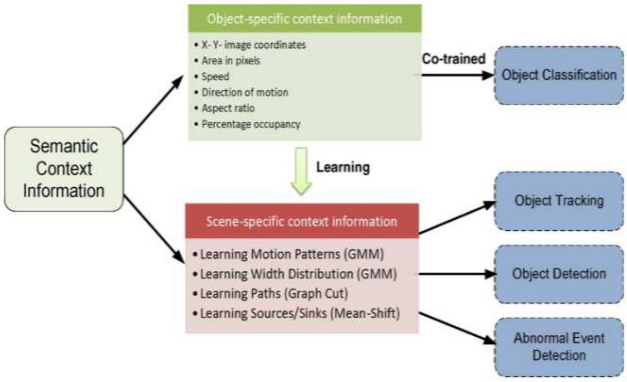

Fig. 6 Mining semantic context details architecture [8]

In this approach, moving objects will be tracked and identified by implementing real-time background subtraction and tracking technique. Then, object-specific context information will be extracted from each foreground object by utilizing the object detection and tracking technique. Based on trajectory analysis, object motion pattern and width distribution will be examined by employing the Gaussian mixture model. Similar motion patterns from the previous process will be grouped together via a graph cut algorithm to obtain paths. Next, the trajectories will be further clustered. A mean-shift-based multiple data mode-seeking algorithm will be employed to analyze the object's entry and exit points and its main trajectories. Finally, object classification, detection, tracking, and abnormal event detection will be enhanced based on the information that has been extracted [8].

1) Mining Scene-Specific Context Details: As established by Zhang *et al.* [8], scene-specific context details expressed the features of objects in the image scene, and it can be studied from the indefinite period of monitorization to differentiate the objects. In this method, scene-specific context details will be learned from each block rather than from each pixel due to the problem that adjacent pixels in the image might have a resemblance in scene context properties. The size of each block is rather tiny. Therefore, it can be assumed that the size of a moving object in certain blocks and its motion pattern are fixed.

There are four (4) fundamental scene-specific context details considered in this method: object motion patterns, widths, paths, and sources or sinks [8]. According to Zhang *et al* [8] object motion patterns can be acquired by harvesting motion direction in each block through its trajectory's inspection whereas distribution of width in individual block can be studied from vehicle's width by protruding its binary mask vertically to the motion direction in respective block. On the other hand, paths and sources or sinks, can also be analyzed according to the motion pattern that have been obtained before.

- Motion Pattern for Each Block.

Zhang *et al* [8] indicated that trajectory can be derived by monitoring the object centroid. There are two types of trajectories which are a trajectory that belongs to human and trajectory that belongs to the vehicle. Motion patterns for each of the trajectory of respective block can be represented via Gaussian distributions from a statistic perspective [8]. In addition, multiple Gaussian models will have to be

implemented because each block may accommodate countless motion patterns. There are four leverages of implementing a Gaussian mixture model to examine motion patterns. First, the computational cost is low. Second, several Gaussian models are adequate to define every single block even though it may accommodate numerous motion patterns since the tight number of traffic rules constrains the number of motion patterns in each small block. Third, by simply updating the Gaussian model's weight, outlier trajectories can be eliminated so that primary motion patterns can be studied from indefinite monitorization. Lastly, the Gaussian model's weight reflects the significance of its corresponding motion pattern that ensures the quantity of major activities can be acknowledged.

- Width Distribution for Each Block.

Each block's width distribution can only be derived after motion direction for each block is extracted according to the motion pattern that has been learned. The width distribution for the block is studied by utilizing the foreground width [8]. However, the width may have distinct differences since the foreground in the traffic scene could either be a single-vehicle blob or a multi-vehicle blob. Thus, the width's probabilistic distribution will be represented as a Gaussian mixture model in every block [8]. Every individual Gaussian module will be considered as one of the fundamental width distributions. After that, Gaussian mixture model parameters will be enhanced with adaptive weights similar to the process of studying each block's motion patterns. Finally, parameters from the Gaussian component's largest weight will be specified as the features for every block.

- Paths for Scene.

A path can be obtained by grouping similar motion pattern with the help of a clustering algorithm [8]. However, Zhang *et al.* [8] further asserted that spatial relations between local blocks would not be considered by some algorithms such as K-means algorithm. It is crucial to take account of spatial relations while grouping the motion pattern together because two neighboring blocks may contain similar motion patterns. Therefore, a graph-based algorithm is adopted in this process to solve the problem [8]. The graph-cut algorithm in this method is employed to acquire the respective semantic region for each motion pattern. After all the semantic regions are derived, trajectories that match the similar semantic regions will be considered as a cluster. After that, trajectory analysis will be employed for each cluster to elongate the trajectories distribution in order to extract its corresponding paths. Finally, a mean-shift algorithm will be adopted to acquire the primary trajectory.

- Sources/Sinks.

Sources and sinks are the positions in which the vehicles enter or leave the scene. False entry and exit might transpire during the estimation of sources and sinks of the vehicles. Therefore, the mean-shift algorithm is the most suitable technique to find these sources and sinks points for each trajectory cluster to ensure they are the boundaries of the path regions [8].

2) Application of Semantic Context Details:

- Enhancement of Object Classification.

Labelling a huge amount of training set manually to train the classifier is inefficient and tiresome. Not to mention, obtaining an enormous set of labeled samples to train the two classifiers is expensive and costly. Therefore, inspired by co-training learning, a semi-supervised method is implemented to study the two classifiers which are AdaBoost classifier and LDA-based classifier [8]. AdaBoost classifier is actually a composition of multiple weak classifiers that group together in order to become strong and better classifiers [13]. Two sets of features that consist of object-specific context features and appearance features based on MB-LBP will also be defined [8]. Based on these features, two labelled sets are then assembled to be used for training each classifiers. Finally, each classifier guesses the unlabeled samples to widen the other's training set.

- Enhancement of Object Detection.

Based on the scene-specific context information that has been extracted before, two steps need to be followed to improve object detection in an intelligent traffic surveillance system [8]. The first step is to categorize the foreground into single-vehicle or multi-vehicle objects by implementing a classifier. In this case, the classifier that has been selected is the Bayes classifier. Zhang *et al.* [8] explained that it was chosen because it is useful in determining whether the foreground belongs to single-vehicle (SV) or multi-vehicle (MV). Next, the subsequent step that needs to be done is partitioning the multi-vehicle blob into a single-vehicle blob. Vehicles in a blob may possess the same texture, color, and shape feature, making it quite problematic to segment a blob into a single vehicle based on these characteristics. Nevertheless, scene-specific context features such as motion direction and width distribution of vehicles are secure in a fixed scene. Therefore, it can be used to assist the segmentation of multi-vehicle blob. As a result, an original approach formulated on scene-specific context features will be employed to increase vehicle identification precision.

Enhancement of Object Tracking. Motion pattern that has been studied before can be used to predict an object's motion by means of trajectory [8]. The initial part of a motion trajectory with k points needs to be specified to compute its probability under each motion pattern to signify the probability that the object is predicted to move along the trajectory portrayed by the motion pattern. The maximum probability will be selected as the most potential one along which the object is predicted to move. However, if the probability is inadequate, it will be discarded from becoming a prospective trajectory for the object. On the other hand, the motion pattern that hold the highest likelihood with the trajectory will be implemented to boost object tracking.

- Abnormal Event Detection.

Abnormal events can also be detected by learning the motion patterns. It signifies a violation of traffic rule by the vehicles by means of analyzing their trajectories. Therefore, Zhang *et al.* [8] emphasized that the probability of trajectory under each motion pattern will be computed to search for the motion pattern that holds the highest likelihood with the trajectory. If the probability of trajectory under the motion pattern is smaller than a threshold, the trajectory will be considered as abnormal. A threshold is calculated by utilizing the probability of each trajectory that corresponds to the given motion pattern.

3) Experimental Analysis

Object Classification: First, for object classification result, three (3) classifiers which are LLC classifier, AdaBoost classifier and LDA-based classifier are compared with the proposed co-training classifier in six (6) scenes. The AdaBoost classifier is trained with 41934 negative samples (vehicles) and 20213 positive samples (pedestrians) labelled manually whereas the LDA-based classifier and LLC classifier are trained with 3500 negative samples and 12000 positive samples for each scene using scene context features. LLC classifier is a locality-constrained linear coding approach for image classification, which implement locality constraints to project each descriptor into its local-coordinate system. After that, the projected coordinates are integrated by maximum pooling to generate a final representation in which a linear SVM classifier is trained. As for the proposed co-training classifier, training samples are initialized with 6716 negative samples and 2720 positive samples labeled manually. Even though a huge set of labeled training samples is not adopted in the proposed co-training classifier, it still accomplishes more notable performance and achieves higher percentage accuracy even in a diverse scene. Table I shows the detailed classification result.

TABLE I
CLASSIFICATION RESULTS OF THE FOUR CLASSIFIERS [8]

Scene	S1%	S2%	S3%	S4%	S5%	S6%
LLC classifier	89.4	90.6	90.2	93.5	87.6	88.7
LDA-based Classifier	87.1	88.3	91.3	91.5	82.5	84.1
AdaBoost Classifier	91.1	87.3	89.8	90.3	80.5	85.3
This Study	98.2	97.3	96.6	97.4	96.8	97.8

Object Detection: 2138 vehicle sequence has been collected from ten (10) various scenes to validate the classifiers' performance. The authors employed Bayes classifier to group each blob and separate the multiple-vehicle blobs into several single-vehicle blobs. According to the result, the method achieves satisfactory performance in various scenes. Table II demonstrates the result of the classification. Once the object is labeled as a multiple-vehicle blob, a segmentation module will be deployed. The algorithm is evaluated in eight different (8) scenes where 11365 multiple-vehicle blobs have been detected, and 10797 of them have been properly partitioned into single-vehicle blobs. The successful partition rate is about 95.0%. Fig. 7 illustrates some results of the segmentation where the red boxes are detected using GMM algorithm, whereas the blue boxes are the segmentation module's outcome.

TABLE II
CLASSIFICATION RESULTS OF THE TEST DATASET [8]

	Tracks	Correct Classification	Correct Rate (%)
M-Vehicle	1382	1279	92.6
S-Vehicle	756	709	93.8

Fig. 7 Some segmentation results in the eight scenes dataset [8]

Object Tracking: Next, to show the result for object tracking, an object's trajectories need to be derived. 213 trajectories have been selected for evaluation. The accuracy rate of detection is 92.5%, where 197 trajectories can be predicted correctly. An example of prediction in an actual traffic scene is demonstrated in Fig. 8. The percentage next to the trajectory indicates the probability in which the car is predicted to travel along the trajectory. Once the car starts to move, these three trajectories also shifted into higher or lower probability until the motion is clear.

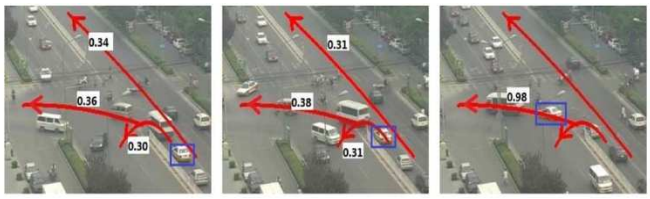

Fig. 8 Motion prediction in actual traffic scene [8]

Abnormal Detection: Finally, to prove the result of abnormal detection, Fig. 9 will be referred. Abnormal detection is divided into two (2) scenes. Rectangle is used to describe boundaries of the six semantic scene models of vehicles for scene S1 as shown in Fig. 9. They are then marked using integer value from one (1) to six (6) in which vehicle in the fifth path will be represented as "RN=5". The lane-merging activity will transpire once a vehicle switches from one path to another. Fig. 9(c) demonstrates the outcome. Other than that, for scene S2, as soon as an object enters the scene, it will be grouped into two (2) options which are pedestrian or vehicle. If the probability of the trajectory is smaller than the threshold, it will be considered as abnormal activities. Among 100 trajectories selected from traffic scene S2, 21 abnormal trajectories and 79 normal trajectories have been successfully detected.

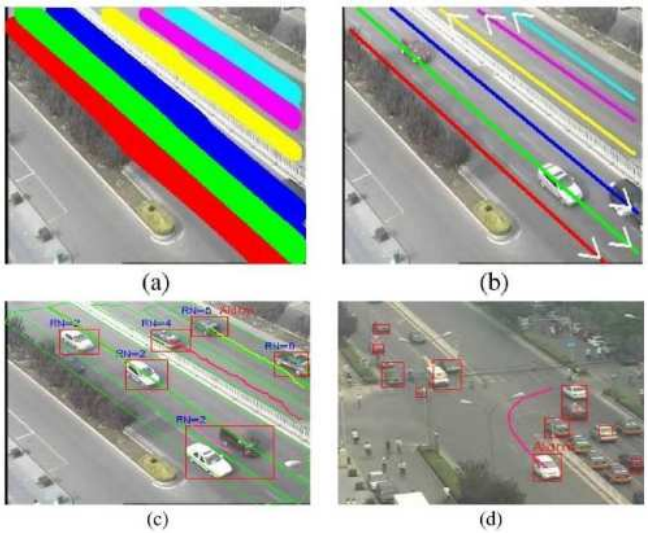

Fig. 9 (a) Semantic scene models represented by six rectangles in scene S1. (b) Semantic scene models main trajectories. (c) Lane-merging occurrence in scene S1. (d) Anomalous trajectory occurrence in scene S2 [8]

C. *Integrating Vision and Language: Semantic Description of Traffic Occurrences*

In this proposed method, case frame analysis is employed to extract traffic occurrences from intersection traffic image sequences that will be implemented in the specialty of natural language processing [9]. Not only this method identified common traffic occurrences; it also spotted abnormal events through knowledge data that describe traffic occurrences as constraining knowledge [9]. The editing approach of knowledge data should be simple and uncomplicated in order to detect various types of traffic occurrences. Therefore, this method suggested a simple text to define knowledge data [9].

Basically, a stationary camera will be fixed upon a pole near the intersection to obtain sequence traffic images. Then, traffic images will undergo a traffic image inspection stage where objects in motion together with their trajectories are identified using an image processing technique. Afterward, traffic events will be extracted through the semantic analysis stage by referring to the knowledge database. If traffic occurrences clash with any of the data enlisted in the knowledge database, it will be specified as an abnormal situation. The following explained the essential processes that are involved in the traffic image analysis stage and semantic analysis stage.

1) Traffic Image Analysis. Traffic image analysis stage will be divided into four (4) main processes as shown below:

- Estimation of Background Image.

The background image is presumed to identify moving objects from each frame image. the Bayesian learning method is suitable to extract background statistics of a dynamic scene [10]. Therefore, this proposed method implements a recursive Bayesian learning mechanism capable of estimating the mean and covariance and its mean and covariance probability distribution for each model [9].

- Tracking Moving Objects.

During this stage, an object tracking algorithm for low-frame-rate video where objects have fast motion is employed [9]. However, this algorithm has a slight issue because the traditional mean-shift tracking might not work if the relocation of an object is huge and the areas between continuous frames do not converge. Nonetheless, Hirano *et al* [9] confirmed that the issue can still be solved by adopting several kernels that are positioned at the center of high motion areas.

- Coordinate Conversion and the estimation of physical parameters.

During this process, domain around intersection on two-dimensional image will be mapped into 3D virtual space that is normalized to 100x100x100. Then, the physical parameters' values of the object in motion at time t will be measured based on the trajectory on the normalized virtual space. There are six (6) physical parameters that have been chosen which are velocity, acceleration, position, the direction of movement, size and distance of two (2) objects [9].

- Detection of Object Types.

Finally, object type will be differentiated and extracted from the normalized size of the moving objects.

2) Semantic Analysis of Traffic Events: Traffic events will be represented using case grammar of natural language processing [9]. According to Hirano *et al* [9], case grammar that has been chosen in this method is suggested by Fillmore where it defines the relationship between a verb and other components which is usually nouns of a single proposition. This method uses case terms (AG, CAG, LOC) and semantic classifications (PHYSOBJ, LIVING, HUM, PHYSLOC) for traffic occurrence identification. There are several benefits of implementing Fillmore's case grammar which are enlisted below:

- Case grammar is the most ideal method to describe the conceptual structure acquired from nonverbal contents since it can express the deep semantic structure in the specialty of natural language processing.
- It simplifies the way to describe and understand traffic events.
- The concept can be conveyed in a simple and concise description.

The knowledge database is produced manually in advanced to extract traffic occurrences from physical parameters and trajectories of objects in motions. According to [9], it comprises four types of data that classify the following contents:

- Object data to specify the object name, attributes such as size or area, and semantic category.
- Physical parameter definition to store the list of physical parameters that have been gathered in image inspection process.
- Predicate verb definition to define the relationship between predicate verb and physical parameters.
- Case frame data to define traffic occurrence.

As stated before, this method considers knowledge databases as constraining knowledge. Therefore, any occurrence that does not satisfy the constraining knowledge data will be considered an abnormal situation [9]. In this method, any extracted occurrence that has a value $1<E<0$ will be categorized as an abrupt abnormal event.

3) Experimental Analysis

Virtual Traffic Sequence Simulation: The algorithm is employed in three settings: mock trajectories that simulate common traffic situations and abnormal circumstances that are usually difficult to obtain in actual traffic environments. Fig. 10 shows the event extraction result that also includes abnormal circumstances. According to the results, all common events were spotted precisely. In addition, abnormal circumstances that were already predicted, such as "A bicycle collides with a bicycle on the footway," were accurately recognized by the algorithm as well as unpredicted circumstances such as "A car runs on the unidentified place".

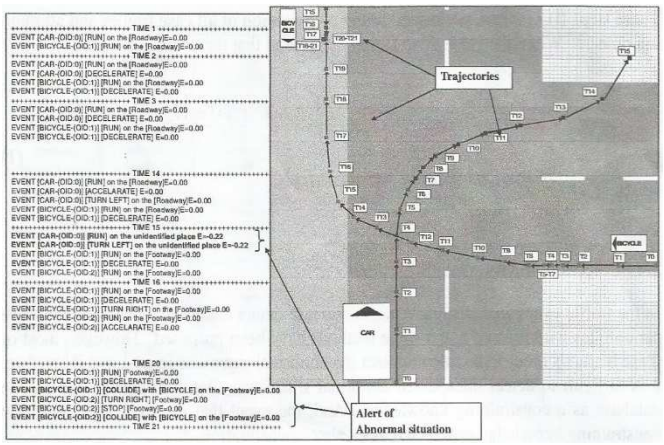

Fig. 10 Traffic extraction result from virtual traffic sequence simulation [9]

Real Intersection Traffic Image Sequence Results: A six (6) minutes traffic video consisting of 5409 image frames is derived from stationary camera assembled on a footbridge that roughly around 5 meter in height. The specifications and settings of the camera is shown in Table III. This traffic image sequence entails 81 vehicles ("CAR" which represent car, lorry or van whereas "BICYCLE" which represent bicycle or motorcycle) and 5 pedestrians.

TABLE III
SPECIFICATIONS AND SETTINGS OF IMAGE ACQUISITION SYSTEM [9]

Camera device	DCR-TRV900 (SONY)
Scanning System	Progressive scan
Mode	Automatic (Brightness, Shutter speed, white balance), Zoom: Off
File format	AVI (DVI Compression), 15 frames/sec
Size/Color	720x480 pixels, 24bit color

Based on the extraction result from the intersection traffic image sequence, this method can extract traffic events with 73.8% precision as shown in Table IV. Even though there is 26.2% traffic event extraction error, 15.2% is actually caused by moving object extraction error generated during the image analysis stage. The other 11.0% event extraction error transpired because of the sudden traffic signal change, analyzing a number of cars as one individual object and segmenting one car into two objects. Accurate traffic event extraction results are shown in Fig. 11(a), while traffic event extraction error is shown in Fig. 11(b).

TABLE IV
TRAFFIC EVENTS EXTRACTION RESULTS [9]

Correct	73.8%	15.2%: Moving object extraction error in the image analysis
Error	26.2%	11.0%: Verb estimation error in the semantic analysis. (e.g. [PREDESTRIAL] [RUN]->PREDESTRIAN] [STOP])

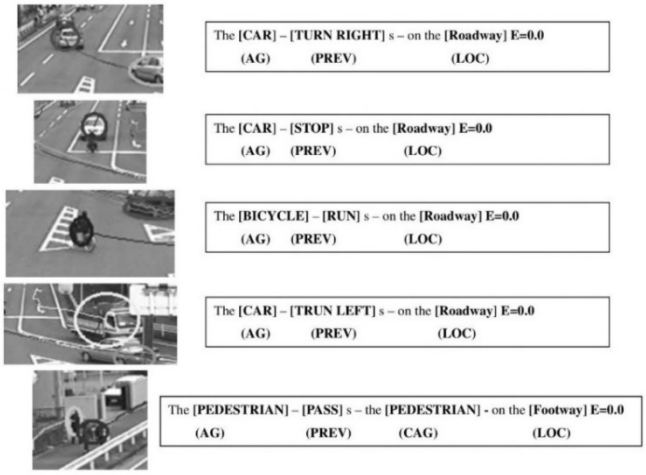

Fig. 11(a) Accurate traffic event extraction results [9]

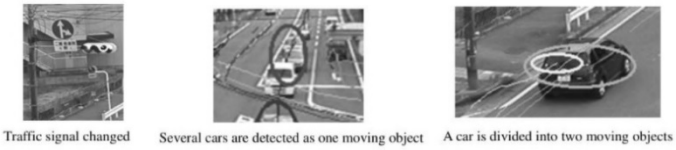

Fig. 11(b) Traffic event extraction error [9]

III. RESULT AND DISCUSSION

There are some issues that need to be resolved in the reviewed approaches above to provide a more robust and efficient method to analyze traffic images semantically. The first issue related to the first approach (semantical analysis of traffic images based on image understanding) is that it can only capture the vehicles' semantic information such as type and color. This approach covers how to extract the semantic information from a vehicle rather than on the semantical approach to extract traffic events [7]. This approach does not provide a technique to detect and extract abnormal traffic events from traffic images. Garg *et al*. [15] stated that traffic event extraction analysis is crucial to provide a more holistic understanding of the monitored traffic area.

The second issue is associated with the fact that the scene-specific context information-based method that has been proposed in the second approach (semantical analysis based on mining semantic context information) is only ideal for medium-field and far-field surveillance videos and only capable of obtaining a better result in this type of traffic scene [8]. This is proven when the segmentation algorithm in this approach can only segment 418 blobs into a single-vehicle blob correctly from 576 multi-vehicle blobs in i-LIDS parked vehicle detection datasets. According to Zhang *et al.* [8], the success rate is only 72.6% which is far lower than the success rate of blob segmentation that they have evaluated in the previous eight (8) scenes dataset which is a 95.0% correct rate.

Finally, the last issue is regarding the drawback in which the error during foreground extraction in the image analysis stage and verb estimation in semantic analysis stage might contribute to the incorrect abnormal events detection in the last approach reviewed in this paper (semantical analysis based on integrating vision and language to describe traffic occurrence). On account of the error accumulated in image analysis stage, there is 15.2% error rate during traffic event extraction process [9]. There is only 73.8% correct traffic event extraction rate from the experiment that has been conducted. Not to mention, the success rate is also affected by 11.0% verb estimation error in the semantic analysis stage [9]. The approach must be improved so that the foreground object misclassification and verb estimation error will not have such hugely influence on the accuracy rate of the traffic event extraction stage.

IV. Conclusion

Since traffic congestion problem has exacerbated over the years, traffic surveillance systems have played an important role in traffic control and management throughout the big cities. Consequently, the existing approach that only extracts the traffic images of their objects and their simple relationship is not sufficient to deliver a traffic management system that is efficient and reliable. Traffic surveillance system should be able to understand and successfully analyze traffic images as if it is able to imitate human perception without human interference.

Therefore, this paper summarizes three semantical analysis approaches to extract traffic density based on traffic images. Some approach understanding technique or attempted to mine the semantic context information from the traffic images sequence whereas others integrate vision and language to semantic Even though these approaches successfully extract traffic images employing semantical analysis, there are still some issues that need to be resolved to increase its precision to obtain a better result.

Acknowledgment

Universiti Malaysia Sarawak funded this research under F08/SpFRGS/1528/2017.

References

[1] A. Joshi and D. Mishra, "Review of traffic density analysis techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 7, pp. 209-213, July 2015.

[2] B. Narang and P. Kochar, "Real time traffic light controller," *International Journal of Computer Technology and Applications,* vol. 5, no. 3, pp. 1092-1096, May 2014.

[3] U. Nagaraj, J. Rathod, P. Patil, S. Thakur, and U. Sharma, "Traffic jam detection using image processing," *International Journal of Engineering Research and Applications (IJERA),* vol. 3, no. 2, pp. 1087-1091, March 2013.

[4] V. Dangi, A. Parab, K. Pawar, and S. S. Rathod, "Image processing based intelligent traffic controller," *Undergraduate Academic Research Journal (UARJ),* vol. 1, no. 1, pp. 1-17, 2012.

[5] N. Lende and S. S. Paygude, "Survey on traffic monitoring system using image processing," *International Journal of Advanced Reseacrh in Computer Engineering & Technology (IJARCET)*, vol. 3, no. 12, pp. 4374-4377, Dec. 2014.

[6] H. H. Wang, D. Mohammad, and N. A. Ismail, "An efficient parameters selection for object recognition based colour features in traffic image retrieval," *The International Arab Journal of Information Technology,* vol. 11, no. 3, pp. 308-314, May 2014.

[7] J. Wu, Z. Cui, H. Yue, and G. Zhang, "Semantic analysis of traffic video using image understanding," *Journal Of Multimedia,* vol. 7*, no.* 1, pp. 41-48, Feb. 2012.

[8] T. Zhang, S. Liu, C. Xu, and H. Lu, "Mining semantic context information for intelligent video surveillance of traffic scenes," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 1, pp. 149-160, Feb. 2013.

[9] T. Hirano, S. Yoneyama, Y. Okada, and Y. Kosugi, "Integrating Vision and Language: Semantic Description of Traffic Events from Image Sequences," *International Symposium on Visual Computing*, 2007, pp. 459-468.

[10] O. Tozel, F. Porikli, and P. Meer, "A Bayesian Approach to Background Modeling," *IEEE Workshop on Machine Vision for Intelligent Vehicles (MVIV)*, vol. 3, pp. 58-65, June 2005.

[11] S. S. Harsha and Ch. Sandeep, "Real time traffic density and vehicle count using image processing technique," *International Journal of Research in Computer and Communication Technology (IJRCCT),* vol. 4, no. 8, pp. 594-598, Aug. 2015.

[12] H. Haeikki, S. Y. Fatemeh, and Chen, K., "Car type recognition with deep neural networks," *IEEE Intelligent Vehicles Symposium (IV)*, pp. 1115-1120, June 2016.

[13] Y. Tang, C. Zhang, R. Gu, and P. Li, "Vehicle detection and recognition for intelligent traffic surveillance system," *Multimedia Tools and Applications,* vol. 76*, no.* 4, pp. 5817-5832, Feb. 2017.

[14] B. Sharma, V. K. Katiyar, A. K. Gupta and A. Singh, "The automated vehicle detection of highway traffic images by differential morphological profile," *Journal Of Transportation Technologies,* vol. 4*, no.* 2, pp. 150-156, Apr. 2014.

[15] K. Garg, S. Lam, T. Srikanthan, and V. Agarwal "Real-time road traffic density estimation using block variance," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, 2016, pp. 1-9..

[16] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance- A survey," *2015 International Conference on Pervasive Computing (ICPC)*, Pune, 2015, pp. 1-6.

[17] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247-2256, Aug. 2015.

[18] S. Qu, Y. Xi, and S. Ding, "Image caption description of traffic scene based on deep learning," *Journal of Northwestern Polytechnical University*, vol. 36, no. 3, pp. 522-526, June 2018.

[19] N. Abid, T. Ouni and M. Abid, "Vehicle detection for intelligent traffic surveillance system," *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Sousse, Tunisia, 2020, pp. 1-5.

[20] N. Audebert, B. L. Saux, and S. Lefevre, "Segment-before-detect vehicle detection and classification through semantic segmentation of aerial images,"*, Remote Sensing,* vol. 9*,* no. 4, pp. 1-18, Apr. 2017.

[21] Y. Tian, Y. Wang, R. Song and H. Song, "Accurate vehicle detection and counting algorithm for traffic data collection," *2015 International Conference on Connected Vehicles and Expo (ICCVE)*, Shenzhen, 2015, pp. 285-290.

[22] L. Suhao, L. Jinzhao, L. Guoquan, B. Tong, W. Huiqian and P. Yu, "Vehicle type detection based on deep learning in traffic scene," *Procedia Computer Science,* vol. 131, pp. 564-572, Jan. 2018.

[23] Y. Zhou and N. Cheung, "Vehicle classification using transferable deep neural network features," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 7, pp. 209-213, Jan. 2016.

[24] V. Vijayaraghavan and M. Laavanya, "Vehicle classification and detection using deep learning," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 1S5, pp. 24-28, Dec. 2019.

[25] S. Kul, S. Eken and A. Sayar, "A concise review on vehicle detection and classification," *2017 International Conference on Engineering and Technology (ICET)*, vol. 4, no. 7, pp. 1-4, Aug. 2017.

[26] K. V. Sakhare, T. Tewari and V. Vyas, "Review of vehicle detection systems in advanced driver assistant systems," *Archives of Computational Methods in Engineering,* vol. 27, pp. 591-610, March 2019.

[27] W. Li and H. Dai, "Real-time road congestion detection based on image texture analysis," *Procedia Engineering,* vol. 137, pp. 196-201, Dec. 2016.

[28] Y. Wang, X. Ban, H. Wang, D. Wu, H. Wang, S. Yang, S. Liu and J. Lai, "Detection and classification of moving vehicle from video using Multiple Spatio-Temporal Features," *IEEE Access*, vol. 7, pp. 80287-80299, June 2019.

[29] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle detection techniques for collision avoidance systems," *IEEE Transactions on Intelligent Transportation Systems,* vol. 16, no. 5, pp. 2318-2338, Oct. 2015.

[30] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images" *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1938-1942, Sep. 2015.

[31] D. Sahgal, Dr. A. Ramesh, and Pof. M. Parida, "Real-time vehicle queue detection at urban traffic intersection using image processing," *International Journal of Engineering Science and Generic Research (IJESAR),* vol. 4, no. 2, pp. 12-15, Apr. 2018.

[32] X. Wen, L. Shao, W. Fang, and Y. Xue, "Efficient feature selection and classification for vehicle detection," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 25, no. 3, pp. 508-517, March 2015.

[33] M. S. Uddin, A. K. Das, and M. A. Taleb, "Real-time area-based traffic density estimation by image processing for traffic signal control system: Bangladesh perspective," *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT),* Dhaka, pp. 1-5, 2015.

[34] R. Takeuchi, K. Kato, D. Harwood, and L. S. Davis, "Vehicle detection using PLS Hough transform," *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision,* Mokpo, 2015, pp. 1-6.

[35] A. Bartsch, F. Fitzek, and R. Rasshofer, "Padesterian recognition using automotive radar sensors", *Advances in Radio Sciences.* vol. 10, pp. 45-55, Sep. 2012.

[36] V. Keerthi Kiran, P., Parida, and S. Dash, "Vehicle detection and classification: a review," *10th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA)*, Odisha, India, 2019, pp. 45-56.

[37] A. Arinaldi, J. A. Pradana, and A. A, Gurusinga "Detection and classification of vehicles for traffic video analytics," *Procedia Computer Science,* vol. 144, pp. 259-268, 2018.

[38] C. Tsai, C. Tseng, H. Tang, and J. Guo, "Vehicle detection and classification based on Deep Neural Network for intelligent transportation applications," *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honululu, HI, USA, 2018, pp. 1605-1608.