This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Hierarchical and Networked Vehicle Surveillance in ITS: A Survey

Bin Tian, Brendan Tran Morris, Ming Tang, *Member, IEEE*, Yuqiang Liu, Yanjie Yao, Chao Gou, Dayong Shen, and Shaohu Tang

***Abstract*—Traffic surveillance has become an important topic in intelligent transportation systems (ITSs), which is aimed at monitoring and managing traffic flow. With the progress in computer vision, video-based surveillance systems have made great advances on traffic surveillance in ITSs. However, the performance of most existing surveillance systems is susceptible to challenging complex traffic scenes (e.g., object occlusion, pose variation, and cluttered background). Moreover, existing related research is mainly on a single video sensor node, which is incapable of addressing the surveillance of traffic road networks. Accordingly, we present a review of the literature on the video-based vehicle surveillance systems in ITSs. We analyze the existing challenges in video-based surveillance systems for the vehicle and present a general architecture for video surveillance systems, i.e., the hierarchical and networked vehicle surveillance, to survey the different existing and potential techniques. Then, different methods are reviewed and discussed with respect to each module. Applications and future developments are discussed to provide future needs of ITS services.**

***Index Terms*—Behavior understanding, computer vision, networked surveillance system, traffic surveillance, vehicle detection, vehicle tracking.**

I. Introduction

WITH the rapid development of urbanization, traffic congestion, incidents, and violations pose great challenges on traffic management systems in most large and medium-sized cities. Consequently, research on active traffic surveillance, which aims to monitor and manage traffic flow, has attracted much attention. With the progress in computer vision, the video camera has become a promising and low-cost sensor for traffic surveillance. Over the last 30 years, video-based surveillance systems have been a key part of intelligent transportation systems (ITSs). These systems capture vehicles' visual appearance and extract more information about them through vehicle detection, tracking, recognition, behavior analysis, and so forth. Generally, existing surveillance systems collect traffic flow information that mainly includes traffic parameters and traffic incident detection. Traffic incident detection is more challenging and has much research potential.

Although great progress has been made on video-based traffic surveillance, researchers are still facing various difficulties and challenges for practical ITS applications. A summary of the existing challenges for video-based surveillance systems are as follows.

- *All-day surveillance*: The lighting conditions changes at different time of a day, particularly between daytime and nighttime. Supplemental lighting equipment can be used for nighttime operation; however, their visual ranges are usually limited.
- *Vehicle occlusion*: In busy traffic scenarios, vehicles are easily occluded by other vehicles and nonvehicle objects, such as pedestrians, bicycles, trees, and buildings.
- *Pose variation*: Vehicle pose can vary greatly when they are turning or changing lanes.
- *Different types of vehicles*: There are various vehicles with different shapes, sizes, and colors.
- *Different resolutions*: When a vehicle drives through the camera's field of view (FOV), its image size in pixels changes. This leads to the loss of some detailed visual information and challenges the robustness of detection models.
- *Vehicle behavior understanding on the road network*: Tracking a vehicle traveling through the road network requires the cameras coordinate with each other in order to understand the full traffic status through global behavior analysis.

Based on the analysis of existing surveillance systems, we present a general system architecture of hierarchical and networked vehicle surveillance (HNVS) in ITSs (see Fig. 1) with the aim of vehicle attribute extraction and behavior understanding. The HNVS hierarchy is constructed from four layers, which are defined as follows.

Layer 1 *Image Acquisition*: The function of this layer is to sense traffic scenes and obtain images using visual sensors.

Manuscript received February 15, 2014; revised May 31, 2014; accepted July 9, 2014. This work was supported in part by the National Natural Science of China under Grant 71232006, Grant 61233001, Grant 61101220, and Grant 61375035 and in part by the Ministry of Transport of China under Grant 2012-364-X18-112. The Associate Editor for this paper was F.-Y. Wang.

B. Tian, Y. Liu, Y. Yao, and C. Gou are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: bin.tian@ia.ac.cn; yuqiang.liu@ia.ac.cn; yanjie.yao@ia.ac.cn; gouchao@casc.ac.cn).

B. T. Morris is with the Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, Las Vegas, NV 89154-4026 USA (e-mail: brendan.morris@unlv.edu).

M. Tang is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tangm@nlpr.ia.ac.cn).

D. Shen is with the College of Information System and Management, National University of Defense Technology, Changsha 410073, China (e-mail: dayong.shen89@gmail.com).

S. Tang is with the Beijing Key Laboratory of Urban Intelligent Traffic Control Technology, School of Mechanical and Electronic Engineering, North China University of Technology, Beijing 100041, China (e-mail: tshaohu@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2014.2340701

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

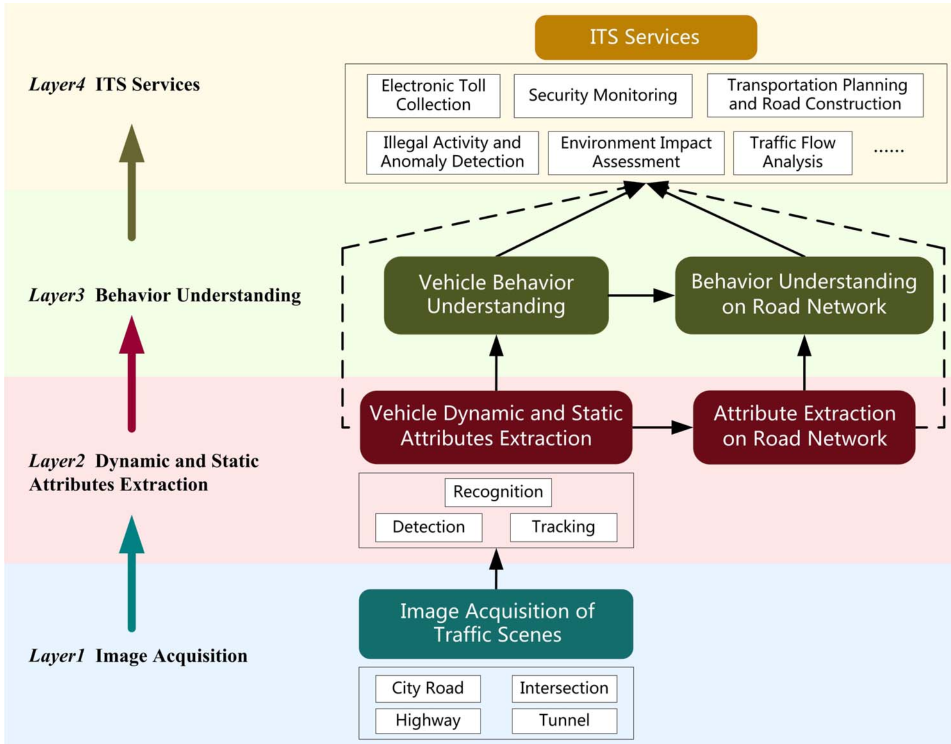

Fig. 1. Architecture of Hierarchical and Networked Vehicle Surveillance.

Layer 2 *Extraction of Dynamic and Static Attributes*: Based on the obtained images, this layer is used to extract the vehicle's dynamic and static attributes. The dynamic attributes refer to the attributes with respect to vehicle motion characteristics, including velocity, direction of movement, vehicle trajectories on a single camera and on the road network, etc. The static attributes represent the features of vehicle appearance description, which consist of license plate number, type, color, logo, etc.

Layer 3 *Behavior Understanding*: This layer aims to analyze the vehicle's dynamic and static attributes, understand vehicle behaviors, and finally perceive traffic status of the transportation system. Vehicle behaviors are analyzed both from a single camera and over the road network in order to obtain and predict the traffic status of the whole transportation system.

Layer 4 *ITS Services*: Based on the outputs of the previous layers, this layer provides ITS services for efficient transportation management and control. Example ITS services include electronic toll collection, security monitoring, illegal activity and anomaly detection, and environment impact assessment.

HNVS is both hierarchical and networked. The functions have little overlap between different layers, which simplifies analysis of layer techniques in literature. Since the HNVS is a networked system, it is possible to generate full networked conclusions, e.g., capturing and understanding the vehicle behaviors on the road network, and perceiving and predicting the traffic status of the whole transportation system.

A recent review [1] presented the state-of-the-art computer vision techniques for the analysis of urban traffic. This survey focused on Layer 2 techniques such as vehicle foreground segmentation, vehicle classification were reviewed in detail, and the complete traffic surveillance system was discussed on both urban and highway environment. Further, the five key computer vision and pattern recognition technologies required for large-area surveillance, including multicamera calibration, computation of the topology of camera views, multicamera tracking, object reidentification and multicamera activity analysis, has been reviewed with detailed descriptions of their technical challenges and comparison of different solutions [2]. Different from [1] and [2], we depart from the typical perspectives of general hierarchical and networked surveillance to consider the problem of vehicle surveillance explicitly. We survey both motion-based and feature-based vehicle detection methods and the latest advances in other fields about surveillance, e.g., vehicle tracking, recognition, and behavior understanding. Furthermore, we present a vehicle surveillance framework for monitoring and understanding their behaviors both on a single camera and the road network. The contribution of this survey is threefold. First, it presents the HNVS architecture to consider the problem of vehicle surveillance from the perspectives of hierarchical and networked surveillance. Second, it provides a comprehensive latest review of state-of-the-art computer vision techniques used in traffic surveillance. Third, we present detailed analysis, discussions, and outlooks on special computer vision issues and surveillance systems.

The remainder of this paper is organized as follows. Section II provides a comprehensive survey of the state-of-the-art research on dynamic and static attributes extraction with detailed discussion with respect to each task. Section III reviews the literature about vehicle behavior understanding, i.e., vehicle behavior understanding both on a single camera and on the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

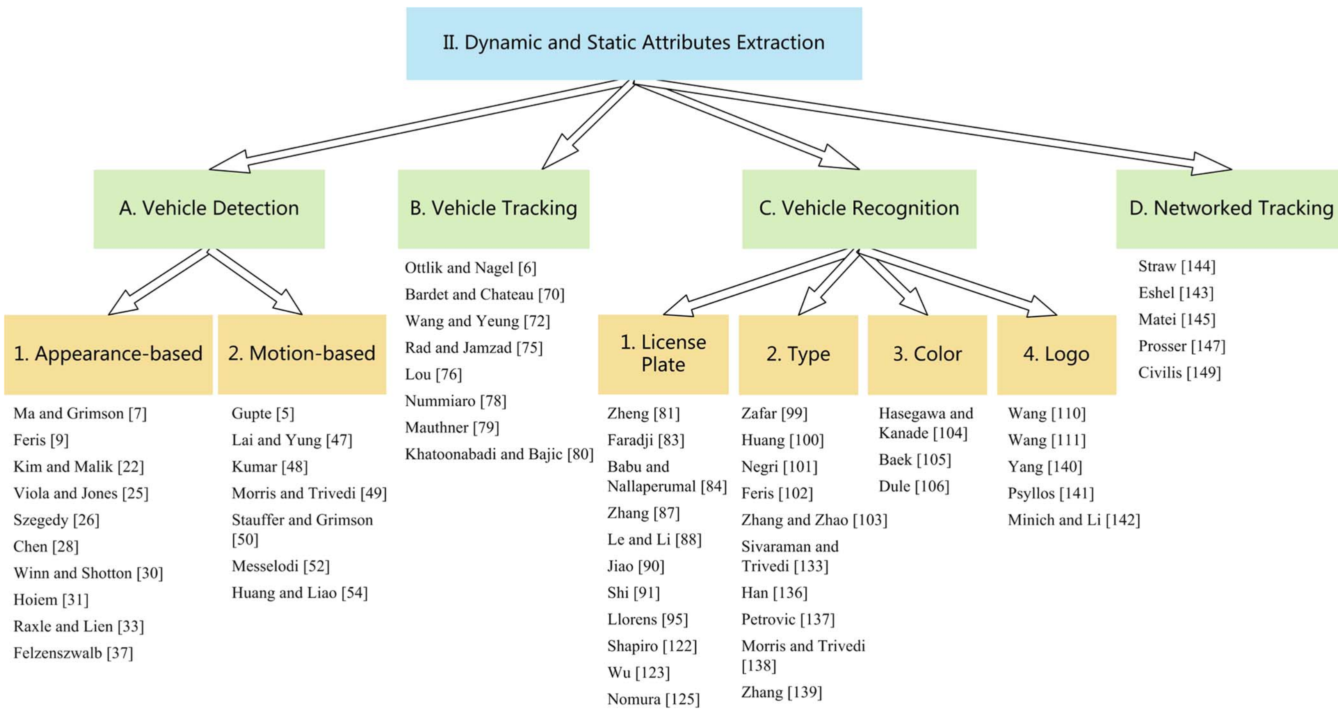

Fig. 2. Taxonomy for vehicle attributes extraction methods and the lists of selected publications corresponding to each category.

road network. In Section IV, the layers of image acquisition and providing ITS services are discussed, and in Section V, the outlook of future developments in video surveillance system are presented. Finally, Section VI summarizes the HNVS framework and concludes this paper.

II. Dynamic and Static Attributes Extraction

Here, we will describe the layer of dynamic and static attributes extraction, review the existing techniques, and provide detailed discussions of challenging issues. Here, the dynamic attributes refer to the attributes with respect to vehicle motion characteristics, including velocity, direction of movement, vehicle trajectories on a single camera and on the road network, etc. It involves techniques of vehicle detection, tracking, and tracking by the camera networks. The static attributes represent the features of vehicle appearance description and recognition, including recognition of license plate number, color, type, logo, and so forth. Fig. 2 demonstrates the taxonomy for vehicle attributes extraction methods and the lists of selected publications corresponding to each category.

A. Vehicle Detection

Reliable and robust vehicle detection, or localization in an image, is the first step of video processing. The accuracy of vehicle detection is of great importance for vehicle tracking, vehicle movement expression, and behavior understanding and is the basis for subsequent processing. There are two main detection research categories.

1) vehicle detection methods based on appearance features;
2) vehicle detection methods based on motion features.

Appearance-based techniques use the appearance features, e.g., shape, color, and texture, of the vehicle to detect the vehicle or separate it from the background. Motion-based methods use "moving" characteristic to distinguish vehicles from the stationary background image. Table I summarizes the representative works in vision-based vehicle detection.

1) Methods Based on Appearance Features: The visual information of object falls into three classes: color, texture, and shape. Prior information is usually employed for modeling when using methods based on these features. In contrast to motion-based methods, appearance-based methods can detect and recognize stationary cars.

a) Representative feature descriptors: Representative feature-based approaches concern methods using coded descriptions about the inherent visual appearance of vehicle, such as symmetry [10]–[12], color [3], [13], edge [14]–[16], contour [17] and texture [18], [19]. A variety of feature descriptors have been used in this field to describe the vehicle's visual appearance.

Many earlier works [20] used local image patches to represent vehicle objects. It is simple to use the patch pixel values as the feature vector, but this representation is sensitive to the vehicle size and illumination changes. As a result, edge-based histograms [7], [21], [22] are used to achieve more spatial invariance and to deal with the influence of illumination conditions.

The scale-invariant feature transform (SIFT) [21] is one of the most widely used local features. This feature is invariant to image scaling and rotation and is partially invariant to illumination change and affine projection by considering local edge orientations around stable keypoints. The modified SIFT descriptor was used in [7] to generate a rich representation of

TABLE I
REPRESENTATIVE WORKS IN VISION-BASED VEHICLE DETECTION

Ref	Motion / Appearance	Description	Comments
Tsai et al. [3]	Appearance	Multi-features: color, corners, edge maps, and wavelet coefficients, radial basis function network, Bayesian and AdaBoost classifiers	A novel color model and a multi-channel classifier
Cucchiara et al. [4]	Motion	Multi-features: color, corners, edge maps, and three-frame difference	Detect vehicles in urban traffic scenes by means of rule-based reasoning on visual data
Gupte et al. [5]	Motion	Multi-features: color, corners, edge maps, and Background subtraction	Method is based on the correspondences between regions and vehicles. Experiment is performed on highway scenes.
Ottlik et al. [6]	Motion & Appearance	Optical flow and edge features, 3D wireframes	Experiments on urban traffic videos with entire automatic initialization
Ma et al. [7]	Appearance	SIFT and edge features, Bayesian classification	Vehicle classification for mid-field video surveillance.
Buch et al. [8]	Appearance	3D histogram of oriented gradients	Extension to HOG feature extraction by applying 3D spatial modelling
Feris et al. [9]	Appearance	Haar-like, AdaBoost	A novel detection/tracking approach for capturing vehicles in challenging urban environments

vehicle images. This histogram of oriented gradient (HOG) [23] is another popular feature descriptor that counts the occurrences of gradient orientation in localized portions of an image. HOG features have illumination invariance and geometric invariance with high computational efficiency on dense sampling grids as opposed to sparse representation in SIFT. Haar-like features [9], [24], [25] compute the difference of the sum of pixels within the rectangles over an image patch. The rectangular structure of Haar-like features is highly efficient to compute, making them well suited for real-time applications and representing rigid objects such as vehicles.

b) Classifiers: Classifiers can be broadly split into two categories, i.e., discriminative and generative classifiers, which follows the general trends in the computer vision and machine learning literature. Discriminate classifiers learn the posterior probability of classification or decision boundary between classes and are more widely used in vehicle detection, whereas the generative classifiers, which learn the underlying distribution or structure of a given class, are less common in the vehicle detection literature.

Discriminative approaches, such as artificial neural networks (ANNs), support vector machines (SVMs), boosting, and conditional random fields (CRFs), are usually preferred for two-class classification: vehicle or nonvehicle. Traditionally, the shortcomings of ANN's were uncertain network topology, many parameters to tune, and potential for locally optimal solutions. However, recent advances in deep learning architectures, such as deep neural networks (DNNs) [26], [27] for general object detection, have been pulling researchers back. In contrast to ANN, SVMs have a much smaller number of tunable parameters and are widely used for vehicle detection [28]. Boosting [29] uses a weighted combination of weak classifiers to create a strong classifier. Feris *et al.* [9] used adaptive boosting (AdaBoost) to deal with a huge pool of local feature descriptors for vehicle detection. Another alternative approach is the use of a probabilistic graphical model, which has been widely used in artificial intelligence, pattern recognition, and computer vision. In [30], CRFs were used to recognize a target. The 3-D Layout CRF was proposed in [31] for joint vehicle detection and pose recognition.

Generative classifiers are not as widely used as discriminative classifiers for vehicle detection. In [32], a spatiotemporal Markov random field model (MRF) was used to achieve robust tracking in occluded and cluttered situations. In [33], Gaussian mixture modeling was used for vehicle detection. In [34], the active basis model, which is a generative model, was proposed and used for object detection and recognition. In this model, a deformable template is represented as an active basis, which consists of a small number of Gabor wavelet elements at selected locations and orientations. These elements are allowed to slightly perturb their locations and orientations to achieve an optimal match with the image.

Part-based detection models, which divide an object into a number of smaller parts and model the spatial relationships between these parts, has become popular for vehicle detection recently. Vehicles were separated into front, side, and rear parts to improve detection performance in occlusion and at the edge of the camera FOV [35]. In [36], objects were broken into their constituent parts, and stochastic attribute graph grammars were used to model the variability of configurations and relationships between these parts. The discriminatively trained deformable part model [37], [38] was used in [39] and [40] for robust vehicle detection.

c) 3-D Modeling: Computer-generated 3-D models of vehicles can be used to detect vehicles by appearance matching [41]–[44]. In [45], existing 3-D vehicle models were analyzed, and it was found that the models were either generic but far too simple to utilize high-resolution imagery or far too complex and limited to specific vehicle instances. To span these two extremes, a deformable vehicle model is constructed with a multiresolution approach to fit various image resolutions. At each resolution, a small number of parameters control the deformation to accurately represent a wide variety of passenger vehicles.

TABLE II
RESULTS OF DIFFERENT METHODS TO DEAL WITH OCCLUSION

Ref	Number of Vehicles		Image Size	Processing Time Per Frame	True Positive Rate	Description
	Total	Occlusion				
Huang et al. [54]	282	30	320×240	-	93.9%	Motion vector
Song et al. [43]	9651	3781	720×480	3 s	94.3%	MCMC, 3D model
Pang et al. [61]	3074	-	320×240	2 s	93.8%	Area ratio, cubical model
Zhang et al. [59]	427	249	320×240	< 0.16 s	94.1%	Silhouette characteristic, motion vector, tracking
Kanhere et al. [60]	2709	-	320×240	32 ms	95.5%	Tracking feature points, motion cues

The main difficulty of 3-D modeling approaches is the question of how to obtain accurate 3-D models. To deal with this problem, a kind of synthetic 3-D model was proposed in [46] to build 3-D representations of object classes. However, this kind of synthetic 3-D model does not always match a real vehicle. Generally, 3-D modeling techniques are limited to only a few vehicle types because it is impossible to generate a model for all vehicles on the road. Even with unlimited models, it is unclear how the unique characteristics of each vehicle type can be extracted and represented for efficient matching. More details about 3-D modeling can be seen in the survey [1].

2) Methods Based on Motion Features: Motion detection is an important task in computer vision. In traffic scenes, the most common characteristic of interest is whether a vehicle is "moving" since it is typically only the moving vehicles that are of interest (traffic counts, safety, etc.). Motion detection aims to separate moving foreground objects from the static background in the image.

Background subtraction methods are the most widely studied and used approach for motion detection. Foreground objects are extracted by calculating the difference by pixel between the current image and a background image [5]. In the simplest case, the background image is constructed by specific known background images, e.g., background averaging method, in which a period of image sequences are averaged to obtain a background model [47]. However, in real traffic scenes, the background is usually changing; therefore, this kind of method is not suitable for dynamic traffic scenes. Thus, the background is constructed without known background image, which make the following assumptions.

- Background is always the most frequently observed in the image sequence.
- The background pixel value is the value that has the maximum appearance time at a steady state.

There are several methods based on the above assumptions, such as the image median, Kalman filter [42], single Gaussian pixel distribution [48], [49], Gaussian mixture model (GMM) [50], [51], and wavelets [52]. Background subtraction methods also have low computational complexity, which makes it suitable for practical applications. However, they are sensitive to the background changes caused by factors such as illumination and weather.

Optical flow is the instantaneous speed of pixels on the image surface, which corresponds to moving objects in 3-D space. The main idea of optical flow is to match pixels between image frames using temporal and gradient information. In [53], dense optical flow was used to separate merged blobs of vehicles. In [6], optical flow was used with 3-D wireframes for vehicle segmentation. The iterative nature of optical flow calculations provides accurate subpixel motion vectors at the expense of added computational time. Yet, optical flow methods are still popular for vehicle detection since these techniques are less susceptible to occlusion issues.

3) Discussion: There are two general problems in the vehicle detection: shadow and occlusion. Here, we will give some discussions about these two problems, as well as challenges with vehicle detection during the nighttime.

a) Vehicle shadow: A shadow is often detected along with the vehicle, particularly when using motion-based methods. Shadow detection is critical for accurate vehicle detection and tracking. In general, a shadow has two features: shape and color/texture. The shape of a shadow has a regular pattern and is determined by the shape of the object and illumination characteristics, which can be exploited to detect the shadow [54]. However, the information of the object shape and the illumination characteristics is difficult to obtain and is usually unstable. The other feature of shadow is its color and texture and is often different from the vehicle. In [55], pixels were analyzed in the hue–saturation–value (HSV) color space to explicitly separate chromaticity (color) and luminosity (brightness) to develop a mathematical formulation for shadow detection, which is not possible in standard RGB color space. An effective shadow-eliminating algorithm based on contour information and color features is developed in [56]. Readers can find more information on shadow detection in [57], which presents an evaluation of moving-shadow detection.

b) Vehicle occlusion: Occlusion is another common problem in detection that arises due to the high density of vehicles and the low camera angles used when monitoring traffic. When occlusion occurs, the appearance of a vehicle is obscured from view either by other vehicles or background objects closer to the camera. Table II lists some experimental results with occlusion handling. In [58], Zhang *et al.* considered that the occlusion is caused by the information lost in the projection of a 3-D scene onto a 2-D image plane, and the key point of handling occlusion is the estimation of the lost information. The proposed framework consists of three levels: the intraframe, interframe, and the tracking level, which are sequentially implemented. In their experiments, 427 vehicles were used, including 249 occlusions, and the accuracy was 94.1%. Kanhere *et al.* [59] presented a method for segmenting

TABLE III
REPRESENTATIVE WORKS IN VISION-BASED VEHICLE TRACKING

Ref	**Methods**	**Description**
Morris et al. [49]	Kalman filter	Use a combination of segmentation and motion information. Experimental results demonstrate robust, real-time vehicle detection, tracking and classification.
Unzueta et al. [69]	Kalman filter, MCMC	A two-step tracking approach, which combines the simplicity of a linear 2-D Kalman filter and the complexity of a 3-D volume estimation using Markov chain.
Bardet et al. [70]	Particle filter, MCMC	Integrate a simple vehicle kinematic model within this tracker allows to estimate the trajectories of a set of vehicles.
Maggio et al. [71]	Particle filter, mean shift	A combination of Particle filter and mean shift is used for object tracking.
Kamijo et al. [32]	Spatio-temporal MRF	The model determines the state of each pixel in an image and its transit, and how such states transit.
Wang et al. [72]	Deep learning	Learn a deep compact image representation for visual tracking using deep learning methods.

and tracking vehicles on highways using a camera that is relatively low to the ground. In their framework, feature points were first extracted, and then the distribution of the points was estimated to recognize occlusion vehicles. The accuracy was over 90%.

c) Nighttime detection: One of the largest obstacles to vision-based vehicle detection systems is performance during the night, which is critical for continuous surveillance. During nighttime operation, cameras are extremely light sensitive and have poor contrast, and reflections of vehicle lights cause serious recognition issues. One solution is to install supplemental illumination equipment (e.g., streetlights or infrared illuminators) to artificially provide visual information similar to daytime operation. However, the range of this type of equipment is limited and may require more expensive camera optics.

Instead, researchers have focused on detecting vehicles based on the limited visual information available with standard cameras at night. Gritsch *et al.* [61] constructed a nighttime classification system, smart eye traffic data sensor, which could distinguish between a car and a truck. The frequency of y-values in the region of interest (ROI) were evaluated, and the x-coordinates parallel with the car's motion were projected. The y-histogram contained two peaks that represented the vehicle headlights, and their distance was used to distinguish between trucks and cars. Robert *et al.* [62], [63] proposed a nighttime detection framework that operated on vehicle lights. Candidate lights were located as bright horizontally aligned blobs in the image. A decision tree was used to determine which candidate blobs corresponded to headlights. Inspired by the work in [63], Wang *et al.* [64] proposed a two-layer nighttime detection method. In the first layer, headlights were detected using the same methods as [63] and, in the second layer, a boosted cascade classifier based on Haar-like features was employed to detect the front of vehicles. Chen *et al.* [65] implemented vision-based multiple vehicle tracking at nighttime. Bright object regions called candidate headlights were segmented using an automatic multilevel thresholding technique [66], and size-ratio, area, and distance constraints were used to filter out the nonheadlights components. Then, they tracked vehicle lights using spatial and temporal features, grouped the vehicle components using motion constraints, and recognized cars and motorbikes. Zhang *et al.* [67] modeled the reflection intensity map, the reflection suppressed map, and image intensity into an MRF model to distinguish light pixels from reflection pixels, which can be used to better represent vehicle and improve vehicle tracking.

B. Vehicle Tracking

Vehicle tracking is used to predict vehicle positions in subsequent frames, match vehicles between adjacent frames, and ultimately obtain the trajectory and location for each frame in the camera FOV of the vehicle. In the HNVS architecture, vehicle-tracking techniques are used to extract vehicles' dynamic attributes, including velocity, direction of movement, and vehicle trajectories. Most vehicle tracking algorithms follow one basic principle: vehicles in two adjacent frames are the same if the spatial distance is small. Table III summarizes the representative works in vision-based vehicle tracking.

1) Vehicle Representation: Before tracking the vehicle, it needs to be uniquely represented. Existing representations utilize the vehicle region, a feature vector, contour, or model. Region representations describe the moving targets with simple geometric shapes, such as a rectangle or oval, which is suitable for rigid objects. In [5], tracking was performed based on motion silhouette overlaps. The feature-based approach is suitable for tracking those targets with small area in the image by compactly representing parts of a vehicle or local areas (see Section II-A1a). Contour representation normally uses a closed curve contour to represent moving objects, and the contour can be continuously updated automatically. It is suitable for complex nonrigid targets. The contour of two vehicles was used in [58] to resolve occlusions by considering the convexity of the shape. Model representation methods generally use 2-D or 3-D appearance models. A 3-D wireframe model was used in [6] and [72] for vehicle tracking. The choice of representation type is closely related to specific applications, the behavior of moving object, or the accuracy requirement.

2) Kalman Filter Tracking: The Kalman filter [73], also known as linear quadratic estimation, is an algorithm that uses a series of measurements observed over time, containing noise and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. The Kalman filter can make full use of the historical information and reduce the search range of the image, to significantly improve system processing speed. Particularly when the vehicle motion and light conditions change or overlapped by other objects, which may cause tracking

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
POPULAR METHODS FOR VEHICLE ATTRIBUTE EXTRACTION

Vehicle Attributes	Features	Classifiers
License Plate Number	Edge information [81] [82] [83] [84] [85] [86] Harr-like [87], Color space [91] [92] [93] [94],	AdaBoost [87] [88] [89], ANN [90], HMM [95], SVM [96]
Type	Edge information [97] [98], oriented-contour [101], PHOG [103]	2DLDA [99] [100], ANN [102]
Color	Color space [104], HSV [105]	k-NN [104], ANN [106], SVM [106]
Logo	SIFT [21], Edge information [110] [111]	ANN [107] [108] [109], Template match [111]

failure, the Kalman filter shows better tracking accuracy and stability. The filter was used successfully in [42], [49], and [74] for tracking. However, the algorithm performance on large motor vehicle tracking is not entirely satisfactory due to linearity assumption and normally distributed noise characteristics. The extended Kalman filter can be utilized to deal with nonlinear models [75].

3) Particle Filter Tracking: The particle filter is a generalization of the Kalman filter. The basic idea of particle filter is to use a set of random samples with associated weights and estimation based on these samples to represent the posterior probability density. According to Monte Carlo theory, when the number of particles is big enough, the group of particles with associated weight can completely describe *a posteriori* probability distribution. At this point, the Bayesian estimation of particle filter is optimal [76]. This approach overcomes the constraint of a single Gaussian distribution of Kalman filters. In [69], [77], and [78], the filter was used for traffic videos. In [70], a combination of particle filter and mean shift was used for object tracking.

4) Dense Inferencing Architectures: The probabilistic graphical model is another mathematical tool for tracking to solve the inference problem of motion estimation. In [32], a spatiotemporal MRF was used to model a tracking problem. The model determines the state of each pixel in an image and its transition, and how such states transit. The spatiotemporal MRF was used in [79] to track moving objects in H.264/AVC-compressed video sequences.

In [71], Wang *et al.* studied the challenging problem of tracking a moving object in a video with possibly very complex background. Inspired by deep learning architectures, they trained a stacked denoising autoencoder using many auxiliary natural images to learn generic image features. This is the first work on applying DNNs to visual tracking, and many opportunities remain open for further research.

C. Vehicle Recognition

In this section, vehicle recognition techniques to extract static attributes, including recognition of license plate number, color, type, and logo, are discussed. Table IV summarizes the popular methods used for vehicle attributes extraction.

1) License Plate: License plate recognition (LPR) is used to extract the vehicle license plate information from an image or a sequence of images. The related algorithms are generally composed of three steps: localization of the license plate region, segmentation of the plate characters, and recognition of the plate characters.

a) License plate localization: This step extracts the regions of license plates in an input image, which directly influences the accuracy of an LPR system. Features, such as edge, texture, color, and hybrid features, are usually utilized to accurately localize the license plate.

The method of extracting the license plate depends on the presence of characters in the license plate, which results in significant change in the grayscale levels between the character pixels and license plate background pixels. The change of the grayscale level results in a high edge density in the scan line [111]. Anagnostopoulos *et al.* [112] proposed a sliding-concentric-window method in which license plates are viewed as irregularities in the texture of the image. Zhang *et al.* [86] employed AdaBoost with Harr-like features for license plate localization since license plates normally have a rectangular shape with a known aspect ratio. Edge detection methods are commonly used to find the rectangles of license plates [80], [83], [85].

Many color-based methods are proposed in the literature for license plate localization. The color combination of the plate and the characters is unique to the plate region [90], [91]. Yohimori *et al.* [113] used a genetic algorithm (GA) to identify the license plate color. Jia *et al.* [114] used Gaussian weighted histogram intersection to overcome the various illumination conditions. Deb and Jo [93] adopted a hue–saturation–intensity model to select a statistical threshold for detecting candidate regions. The mean and standard deviation of hue were used to detect green and yellow license plate pixels. Wan *et al.* [92] proposed a novel method to localize the plate by means of a Color Barycenter Hexagon model that is insensitive to illumination.

In addition to the methods based on a single feature, many research works combined two or more features of the license plate, such as texture and color [115] or rectangular shape with texture and color [116]. Lee *et al.* [117] used the local structure patterns computed from the modified census transform to detect the license plate, and a two-part postprocessing with position-based and color-based methods was adopted.

b) Character segmentation: A wide variety of techniques have been proposed to segment each character after plate localization. The method proposed in [118] used the dimensions of each character for fixed segmentations. Meanwhile, the structure of the Chinese license plate is used to construct a classifier for recognition. As pointed out in [119], the plates in Taiwan are all in the same color distribution, i.e., black characters and white background. The localization and segmentation rates were 97.1% and 96.4%, respectively, in the experiment with 332 different test images. In [120]–[122], the connected pixels in the binary license plate image were labeled to segment the characters. These labeled pixels with the correct same size and aspect ratios are detected as license plate characters. However, it fails to detect the plate characters when there are some joined or broken characters. Similar to plate detection, two or more features can be combined to segment the license plate characters. A morphological thickening algorithm was proposed in [123] to segment by means of locating the related lines for separating the overlapped characters. In [124], an adaptive morphology was proposed to detect fragments and merge them based on the histogram obtained by vertical projection of shape pixels in each column of image matrix.

c) Character recognition: Character recognition has some challenges due to the camera zoom factor that results in various character sizes and sometimes noise. In [125], extracted features were compared with prestored feature vectors to measure the similarity, and they were robust enough to distinguish characters even under distortions. In addition, many classifiers such as ANNs [89], hidden Markov models (HMMs) [94], and SVMs [95] can be used to recognize characters after feature extraction. Some researchers integrate two kinds of classification schemes [126], [127], use multistage classification schemes [128], or a "parallel" combination of multiple classifiers [129], [130].

d) Discussion: Although significant progress of LPR techniques has been made in the last few decades and commercial products exist, there is still plenty of work to be done. A robust system should work effectively under a variety of conditions such as indoors, outdoors, nighttime, or with different colors and complex backgrounds and even when the plates been occluded by dirt or lighting.

A uniform evaluation benchmark needs to be constructed to evaluate the performance of different LPR systems. It is essential that the number and quality of testing examples have a direct effect on the overall LPR performance. Except for common test sets, some regulations should be formulated for performance evaluation, such as how to define the accurate localization of license plate and how to calculate the character recognition rate. (For more research and discussion about LPR, see [131].)

2) Vehicle Type: Recognition of vehicle type, which is an important characteristic of vehicle, has become ever more important in the recent years in ITS [132]. For example, accurate recognition of vehicle types could offer valuable assistance to the police in identifying blacklisted vehicles from a mass of traffic surveillance image database [97] or could be used for emission estimation [133]. The researches of vehicle classification methods are mainly based on the following two aspects: shape features and appearances of vehicles.

a) Shape-based methods: Shape-based methods classify vehicles according to their shape features, e.g., size and silhouette. Lai *et al.* [134] proposed to classify the trucks and cars by longitudinal length. The images should be taken in the driving direction to accurately get the length information. Han *et al.* [135] thought that the most stable image features for vehicle class recognition appear to be image curves associated with 3-D ridges on the vehicle surface. Types of SUV, car, and minibus were recognized and yielded an accurate rate of 88%. Negri *et al.* [100] proposed an oriented-contour point model to represent a vehicle type, which exploits the edges in the four orientations of vehicle front images as features and reported a classification performance of 93.1% by using the classifier based on a voting algorithm and an Euclidean edge distance. Shape-based methods might not yield fine-grained vehicle types as different types of vehicles may be similar in size.

b) Appearance-based methods: Appearance-based methods classify vehicles based on their appearance features, e.g., edge, gradient, and corner. Petrovic *et al.* [136] used Sobel edge response and square mapped gradients together with simple Euclidean measure-based decision module and obtained a good performance on classification of 77 different classes. In [97], Munroe *et al.* detected edge features and used k-means to classify five different vehicle classes. In [98], a 2-D linear discriminant analysis (2DLDA) method was proposed to classify 25 vehicle types. The experiment showed that it yielded accuracy of 91%. Huang *et al.* [99] extracted a ROI relative to the located license plate and also used the 2DLDA method on the gradients of ROI and achieved a high recognition accuracy rate of 94.7% for 20 types. It is not sensitive to colors and illumination due to the feature selection of ROI gradients for 2DLDA. In [49], Morris and Trivedi developed a VECTOR system to classify eight different types of vehicles with over 80% accuracy using simple blob measurements in highway scenes. In their later work [9], they used the term shape-free appearance space to denote the image space of objects with the same aspect ratio and learned a single detector for multiple vehicle types (buses, trucks, and cars) by multiple feature planes (red, green, blue, gradient magnitude, and many others), Zhang *et al.* [137] built a two-stage cascade classifier ensemble with rejection option based on a pyramid HOG and Gabor features extracted from frontal images of vehicles. The experimental results showed it yielded an overall accuracy rate of 98.7% for 21 classes.

c) Discussion: Vision-based vehicle type recognition is still a challenging task due to many issues still to be fully explored. In traffic surveillance videos, even the same vehicle type can be incorrectly classified into different types due to different road environments, illumination variation, complex backgrounds, and different camera views. In addition, many types of vehicles have highly similar appearance, and the number of vehicle types is large and is ever expanding as time goes on.

Future works will be devoted to tackling the challenging aforementioned problems. Moreover, obtained attributes such as color, width, height, velocity, and travel direction can be used to search for specific vehicles in large-scale collections from realistic vehicle surveillance settings. In vision-based vehicle classification, feature representation and classification are two principal issues. Although most studies pursue accuracy in recent years, the reliability and time efficiency will be more important to address for widespread deployment.

3) Vehicle Color: Color is another essential attribute of vehicle that has become more widely used for video surveillance.

Some studies have been developed to recognize the vehicle color. Hasegawa and Kanade [103] classified outdoor vehicle colors into six groups, where each group is composed of similar colors such as black, dark blue, and dark gray. They used a k-nearest neighbor (k-NN)-like classifier with a new linearly related RGB color space. In [104], HSV color space was used for vehicle color recognition. This work used the 2-D histograms of H and S channels as input features for an SVM. It classified the vehicle colors into red, blue, black, white, and yellow. Two regions of interest (the smooth hood piece and semi-front of a vehicle) were used along with three classification methods (k-NN, ANNs, and SVM) and combinations of 16 color space components as distinct feature sets to recognize the vehicle, and yielded an accuracy rate of 83.5% [105].

The vehicles in surveillance scenes are almost always outdoors with various illuminations. The color of vehicle may differ with respect to the illumination and camera viewpoint. Future work will focus on finding the reliable feature representation and robust classification methods for vehicle color recognition.

4) Vehicle Logo: Another important attribute of a vehicle is its logo or emblem, which contains important information about the make and model of a vehicle. Since the vehicle logo cannot be tampered with easily, it plays an elemental role in classification and identification of vehicles. We divide the related works into two parts: logo detection and logo recognition.

a) Logo detection: Accurate logo detection is a critical step in the vehicle logo recognition. A novel approach for vehicle logo detection based on edge detection and morphological filter was proposed in [109]. In [110], Wang *et al.* proposed to conduct fast coarse-to-fine vehicle logo localization, although the accuracy was not high in outdoor environments since it was sensitive to lighting conditions. Yang *et al.* also proposed to take a coarse-to-fine localization step to first find out the position of logo and then localize the logo for details with an average recognition accuracy rate of 98% [138]. In [139], the logo was detected from a frontal-view vehicle image using a license plate detection module to localize the position of vehicle license plate and to know the relationship between plate and logo position.

b) Logo Recognition: The work in [139] adopted the SIFT descriptor to perform vehicle logo recognition. This work claimed that the proposed method effectively used many different views of the database features to describe a detected query feature and simultaneously made the recognition process more robust. The method yielded 91% overall recognition success rate but was time-consuming. Then, in [140], an analysis was made to compare the performance of SIFT operator with Fourier operator for logo recognition. The study concluded that the Fourier operator yielded better performance. The logo recognition is also performed through either neural networks [106]–[108] or template matching [110]. Lee *et al.* [106] built a three-layer neural network trained with texture descriptors for recognition. The method demonstrated a recognition rate of 94% for moving vehicles. In [108], a probabilistic neural network was used for classification of the vehicle logo and obtained an accuracy rate of 87%. Wang *et al.* used template matching and edge-orientation histograms with good results in [110].

c) Discussion: Vehicle logo recognition is essential for vehicle surveillance. Successful recognition mostly depends on the accurate extraction of the small logo area from the original vehicle image. However, different conditions, such as occlusion, illumination change, shadow, and rotation, make vehicle logo recognition still a challenging task, particularly for real-time applications. Future works will be devoted to finding more discriminative feature representations and more robust classification methods for real-time applications.

D. Vehicle Tracking on the Road Network

With the development of the GPS and radio-frequency identification (RFID) techniques, large-scale trajectory data are collected from moving objects on road networks. Networking of cameras deployed at urban intersections and other designated places is conducive for similarly tracking vehicles as they traverse the road network. Based on the vehicle detection and tracking results of each single camera, vehicle tracking on the road network can be achieved with networked cameras, called networked tracking. Here, we treat the results of networked tracking as vehicle's dynamic attributes on the road network.

Existing multicamera techniques identify whether camera views are overlapped or spatially adjacent. Much research has assumed that adjacent camera views have overlap and utilized the spatial proximity of tracks in the overlapping area. As described in [2], tracks of objects observed in different camera views were stitched based on their spatial proximity [141], [142]. In order to track objects across disjoint camera views, appearance cues have been integrated with spatiotemporal reasoning [143], [144]. However, in the case that the cameras that are far in distance and the environments are crowded, the camera network topology is not available, and the object appearance may undergo dramatic changes. Thus, object reidentification is used to match two image regions observed in different camera views and to recognize whether they belong to the same object or not, purely based on the appearance information without spatiotemporal reasoning [145]. More techniques about multicamera tracking can be seen in [2].

Compared with the multicamera tracking techniques in [2], we mainly focus on vehicle tracking techniques on the road network for networked surveillance, i.e., networked tracking.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

From the point of our view, there are two significant differences between networked tracking and multicamera tracking.

1) There might be thousands of cameras deployed on the road network, and it is hard to obtain their topologies.
2) There might be multiple vehicles in the FOV of each camera all the time; therefore, vehicle reidentification is strenuously time-consuming.

Consequently, network-based tracking methods are better suited for use in large road networks for traffic surveillance. Most of these techniques are deployed for GPS-based tracking, but they have the potential to be employed in video-based networked surveillance systems. In [146], three tracking approaches were described based on GPS data, including point-based tracking, vector-based tracking, and segment-based tracking. For point-based tracking, the server represents a moving object's future positions as the most recently reported position. An update is issued by a moving object when its distance to the previously reported position deviates from its current GPS position by a specified threshold. For vector-based tracking, a GPS receiver computes both velocity and heading for the object. It assumes that the object moves linearly and with the constant speed received from the GPS device in the most recent update. For segment-based tracking, this is done by means of map matching. Segment-based tracking predicts a moving object's position according to its speed and the shape of the road on which the object is traveling. In [147], trajectories of moving objects on road networks were characterized by large volumes of updates. A central database stores a representation of each moving object's current position. Each moving object stores the central representation of its position and updates it continuously. Experiments using real GPS logs and a real road network were carried out to verify the performance of the system. To efficiently track an object's trajectory in real time, Lange *et al.* [148] proposed a family of tracking protocols by trading the communication cost and the amount of trajectory data off against the spatial accuracy.

For the cameras deployed on the road network, their locations are known *a priori*, giving precise coordinates. In addition, the intelligent algorithms on the camera can automatically extract the vehicle's locations and attributes. Thus, transferring from the aforementioned GPS-based tracking techniques, all the characteristics of the vehicles (described in Section II-C), particularly the license plate, can be utilized to track vehicles over large areas on the road network.

III. Behavior Understanding

After dynamic and static attributes extraction, vehicle behavior understanding will be performed. We depart from the perspective of networked surveillance to understand vehicle behaviors. In this section, we divide the behavior understanding task into behavior understanding techniques on a single camera and that on the road network.

A. Behavior Understanding on a Single Camera

Behavior understanding may be thought of as the classification of time-varying feature data, i.e., matching an unknown test sequence with a group of labeled reference sequences representing typical or learned behaviors [149]. Simply stated, behavior understanding in traffic surveillance describes the location or speed changing of a vehicle in space and time in the video sequence, e.g., running, cornering, and stopping.

A few recent surveys [150], [151] on behavior understanding were carried out with different taxonomies. Subjects of interest are first detected and tracked to generate motion descriptions, which are then processed to identify actions or interactions. Behaviors are usually recognized by defining a set of templates that represent different classes of behaviors [150]. In the cases where behaviors cannot be represented *a priori*, it is common to use the concept of anomaly, namely a deviation from the predefined behaviors [152]. These surveys [150], [151] focus on human behavior understanding, which is usually categorized into four levels: gesture, action, behavior/activity, and interactions. Gestures are elementary movements of the human body parts, such as waving a hand, stretching an arm, and bending. From these atomic elements, actions are the single-person activity where multiple gestures are temporarily organized in the time domain, for example, running, walking, and jumping [153]. A behavior is the response of a person to internal, external, conscious, or unconscious stimuli [150]. For two or more humans and/or objects doing activity, it is called interaction. Carried/abandon bag, a person stealing a bag from another, and pointing a gun are examples of interactions. (For more information about human behavior understanding, see [150] and [151].) Here, we focus on the techniques of vehicle behavior understanding for traffic surveillance. Vehicle behavior understanding is interlinked with human behavior understanding with respect to the processing techniques. In contrast, vehicle behavior understanding depends on vehicle trajectories and other dynamic attributes, such as velocity and acceleration. We will review the techniques of vehicle behavior understanding with these two aspects in the following.

The fundamental problem of behavior understanding is to learn the reference behavior sequences from training samples and to devise both training and matching methods for coping effectively with small variations of the feature data within each class of motion pattern [154]. There are two main steps in behavior understanding: First, a dictionary of reference behaviors is constructed, and second, it checks if a match can be found in the dictionary for each observation. In traffic surveillance systems, there are a great number of vehicle activities that can be viewed as reference behaviors, such as "stalled or slow motion, speeding or fast motion, heading straight, heading right, and heading left" [155] or "moving down and then turning right on the road, U-turn" [156]. When combined with traffic scene knowledge, these reference behaviors can be applied for two main purposes: explicit event recognition, which means giving a proper semantic interpretations, and anomaly detection, such as traffic event detection (illegal stop vehicles, converse driving, congestion, and crashes [157], [158]) and traffic violation detection (red-light running [159], [160] and illegal lane changing [161]).

From a massive amount of behavior research, there are two main ways to understand vehicle behaviors in traffic surveillance systems. The first one is trying to analyze the motion

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
REPRESENTATIVE WORKS ON BEHAVIOR UNDERSTANDING BY A SINGLE CAMERA

Ref.	**Trajectory Clustering**	**Trajectory Modeling**	**Trajectory Retrieval**	**Behavior**
Trivedi (2005) [159]	K-means clustering	-	-	Illegal stop vehicles, converse driving, congestion, and crashes
Kamijo (2005) [164]	Hierarchy and decision surface	ST-MRF	Congestion; Incident	-
Wang (2006) [165]	Spectral clustering	Gaussian mixture models	-	Anomaly detection: reverse driving; turn left
Jiang (2007) [166]	Dynamic hierarchical clustering (DHC)	HMM	-	Abnormal event
Piciarelli (2008) [167]	Single-class support vector machine (SVM) clustering	Normalized Gaussian kernel	-	Anomaly detection; U-turn
Veeraraghavan (2009) [157]	-	Complete stochastic context-free grammars (SCFGs)	-	Stalled motion; speeding motion; heading straight, right, left; south-north, west-north through intersection motion
Huang (2009) [168]	Particle filter tracking	-	-	Breaking, changing-lane driving and opposite-direction driving
Pucher (2010) [169]	-	-	-	Wrong-way drivers, still standing vehicles, and traffic jams
Kasper (2011) [163]	-	Object-oriented Bayesian networks (OOBNs) with Gaussian distribution	-	Illegal lane changing
Hsieh (2011) [170]	C-means Clustering	Fuzzy self-organized map (F-SOM)	-	Abnormal activities: a car turned left into a parking lot of bicycles; a car stopped at the restricted zone
Hu (2013) [158]	Dirichlet process mixture model (DPMM)	Time-sensitive Dirichlet process mixture model (tDPMM)	DFT coefficient vector sequence	Moving down and then turning right on the road; U-turn

trajectory information of a vehicle, called behavior understanding with trajectory analysis. The other one is trying to analyze the underlying information such as the size, velocity, and direction of the vehicle, called behavior understanding without trajectory here. Table V lists the representative works on behavior understanding by a single camera.

1) Behavior Understanding Based on Trajectory Analysis: Most existing traffic monitoring systems are based on motion trajectory analysis. Trajectory analysis [169] is an important and basic research in behavior analysis and understanding. In traffic video surveillance, learning and analyzing the vehicle trajectory is becoming the main method used to understand vehicle behavior because it is relatively simple to extract and the interpretation is obvious [170].

Trajectory dynamics analysis assumes that change, in particular from motion, is the cue of interest for surveillance. A motion trajectory is obtained by tracking an object from one frame to the next and then linking its positions in consecutive frames. The following describes a common solution framework for vehicle behavior modeling and recognition in traffic monitoring systems using the trajectory-based approach. First, spatiotemporal trajectories are formed, which describe the motion paths of tracked vehicles. Then, characteristic motion patterns are learned, e.g., clustering these trajectories into prototype curves. Finally, by tracking the position within these prototype curves, motion recognition is tackled. To summarize, learning and analyzing trajectories include three basic steps [156]: trajectory clustering, trajectory modeling, and trajectory retrieval. (For complete treatment of behavior analysis using trajectories, the reader is directed to the survey by Morris and Trivedi [169].)

a) Trajectory clustering: The key task here is to determine an appropriate number of trajectory clusters automatically. Inappropriate setting of the number of trajectory clusters may result in inaccurate trajectory clustering, particularly when the number of trajectories is very large. One of the earliest research teams working on behavior analysis in video surveillance is MIT's Artificial Intelligence Laboratory. Stauffer and Grimson [171] learned local trajectory features by vector quantization (VQ) on subimages and learned those features similarities using local cooccurrence measurements to do cluster analysis. Wang *et al.* [163] utilized spectral clustering to complete trajectory clustering, and help to detect and predict anomalies. A fuzzy self-organizing neural network based on k-means was proved to be more efficient than VQ in both speed and accuracy in [172]. Vasquez and Fraichard [173] proposed a rather flexible expectation–maximization algorithm to compute the trajectory similarity, and combined complete-link hierarchical clustering and deterministic annealing pairwise clustering together for trajectory clustering. Piciarelli *et al.* [165] believed that a trajectory should be decomposed into different shared parts, and they proposed a clustering method based on a single-class SVM. Spectral clustering and agglomerative clustering are two most commonly used clustering methods. (See [174] for detail.)

b) Trajectory modeling: Each cluster of trajectories is organized as a trajectory pattern. Trajectory cluster modeling, i.e., trajectory pattern learning, means building a model of trajectories in each cluster according to their statistical distribution, such as a hierarchical Dirichlet process and a Dirichlet process mixture model (DPMM). Usually, the motion trajectory patterns are commonly learned using the HMM [164], fuzzy models [168], and statistical methods [175]. When a new unknown video pattern is incoming, the time-sensitive DPMM can be performed by using known normal events. Hu *et al.* [172] applied a Gaussian distribution function to model the trajectory pattern in the learning phase. The HMM has been used to represent trajectories and time series successfully. Researchers from the Institute of Industrial Science at the University of Tokyo utilized spatiotemporal MRF model to separate occluded vehicles, o learned the movement patterns with a HMM model, and recognized motion behavior. Vasquez and Fraichard [173] made use of growing HMM to realize online adaptive statistical trajectory pattern model that could include new movement patterns. In [176], the modeling of univariate time series by autoregressive models was exemplified. Applications that can benefit from semi-supervised learning of trajectory patterns are demonstrated in [155], which were used to detect "south–north through intersection" or "west–east through intersection" and so on.

c) Trajectory retrieval: In this trajectory analysis step, a user can give a query, such as "find all illegal stop vehicles at 8:00–10:00 in the Southwest road", and matching is performed to return all examples in a traffic surveillance database. The query trajectory is matched to the corresponding trajectory pattern based on the posterior probability estimation. The retrieved videos are ranked by the posterior probabilities. There are two common used algorithms to represent and compare trajectories: string matching algorithm and sketch matching algorithm. As a matter of fact, the string-based method can automatically convert a trajectory to a string and match it using its semantic meanings. Vlachos *et al.* [177] utilized the longest common subsequence (LCSS) algorithm complete string-based matching trajectories by performing a frame-by-frame analysis directly on objects' coordinates. The work in [178] assumed a query by example mechanism according to presented example trajectory and the search system could return a ranked list of most similar items in the data set by a string matching algorithm, whereas the sketch-based method projects a trajectory on a set of basic functions and matches it according to its low-level geometrical features. In [179], a real-time sketch-based similarity calculation method to search millions of images was developed. However, the gap between the user's mind and their specified query can still be large even in such a system. The work in [180] took advantage of the complementary nature of these two methods, and a hybrid method combining the sketch-based scheme and a string-based one together to analyze and index a trajectory with more syntactic meanings was proposed. By utilizing syntactic meanings, most impossible candidates can be filtered out, and at the same time, low-level features can be also used to compare different trajectories. Moreover, their method showed good performance in solving the partial trajectory matching problem.

Based on the above discussion, a great number of successful applications of activity analysis to anomaly detection have been presented in literature. They address both complete and incomplete trajectories in various traffic scenarios, including the detection of illegal U-turns, red-light running, and illegal lane changing.

2) Behavior Understanding Without Trajectory: The other way of behavior understanding is to analyze nontrajectory information such as the size, velocity, direction of the object, or queue length and flow of objects. The main idea is to determine abnormal events according to the sudden changes in velocity, location, and direction [181] of the target or if the value of these attribute of behavior does not meet a predefined threshold rule.

Speed is the estimated velocity of a tracked vehicle converted from the image distance to the actual distance by manual roadway calibration. By velocity monitoring, first, the surveillance system can detect congestion and give the upstream section bypass warning; second, some incidents can be quickly detected from the stopped state, which means traffic accident or a violation behavior. Moreover, vehicle velocity measurements have been used to categorize speeding behavior [182] or highway congestion from stalled vehicles or accidents [162]. Kamijo *et al.* [162] used flow and speed to report highway congestion warnings, and they also showed that congestion is not caused by demand exceeding capacity but of inefficient operation of highways during periods of peak demand. Huang *et al.* [166] utilized velocity, moving direction, and position of the vehicle and recognized vehicle activities, including breaking, changing-lane driving, and opposite-direction driving. Kamijo *et al.* [32] employed an HMM to detect events, including bumping accident, stop and start in tandem, and passing. Pucher *et al.* [167] used both video and audio sensors to detect incidents such as wrong-way drivers, still-standing vehicles, and traffic jams on highways.

B. Behavior Understanding on the Road Network

The FOV of a single camera is finite and limited by traffic scene structures. In order to monitor a wide area, many intelligent multicamera video surveillance systems [183]–[185] have been developed by utilizing video streams from multiple cameras.

Multicamera video surveillance is generally achieved with five key computer vision and pattern recognition technologies, including multicamera calibration, computation of the topology of camera views, multicamera tracking, object reidentification, and multicamera activity analysis [2]. By employing multicamera networks, video surveillance systems can extend their capabilities and improve their robustness. In multicamera surveillance systems, activities in wide areas can be analyzed, the accuracy and robustness of object tracking are improved by fusing data from multiple camera views, and one camera hands over objects to another camera to realize tracking over long distances without break [2]. (For more information about intelligent multicamera surveillance, readers are referred to [2].)

However, most published results on multicamera surveillance are based on small camera networks [2], and they focus on specific object tracking and activity analysis, for example, regular vehicle activity and abnormal motion trajectories [186]. In contrast, networked surveillance cannot only monitor the object behavior but also yield to networked conclusions, e.g., obtaining and predicting the traffic status of the road network, finding interesting regions on the road network, etc.

In addition to behavior understanding on a single camera, we propose to understand the vehicle behaviors from the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

perspective of networked surveillance, i.e., understanding vehicle behaviors on the road network. A large amount of research about trajectory analysis on the network has been developed in the fields of mobile computing and location-based systems (LBSs). However, most of them are based on GPS trajectories and traffic simulation rather than video sensors. GPS-based trajectory analysis can produce networked conclusions about the vehicle object behavior and traffic network, including mining trajectory patterns, predicting vehicle movements, discovering anomalies, and discovering interesting regions.

With the advances of computer vision and network techniques and a growing need for safety and security from the public, tens of thousands of cameras in a city for video surveillance should be networked. Video sensors with intelligent algorithms that are deployed on the road network can be treated as high-end intelligent sensors. They can perceive and yield precise positions, and detailed dynamic and static attributes of the vehicle object, as discussed in Section II-C. Afterward, vehicle behaviors on the road network can be understood, and finally, the traffic status of the whole transportation system is perceived, predicted, and understood.

Compared with GPS-based systems, networked video surveillance systems might play a part in the following aspects.

- *Perception of rich information*. Video surveillance systems can obtain detailed vehicle attributes, as listed in Section II-C, which give more support to the trajectory analysis. In addition, they can yield vehicle queuing length, traffic volume, road occupancy, and other traffic parameters, which can help with efficient traffic control.
- *Low cost*. For GPS-based systems, each vehicle should be equipped with at least one GPS device and typically requires higher penetration rates. However, for video-based systems, cameras might be deployed at key road sections and intersections to monitor the same scale region, which is not contingent on the number of vehicles.
- *Complementarity*. In the case of a vehicle without GPS installed or inoperative, video surveillance systems can be complementary solutions.

In the following, we will review existing research that is related to behavior understanding on the road network. These methods might be not designed for traffic surveillance but they have the potential to be used for networked surveillance.

1) Mining of Trajectory Patterns: Mining of trajectory patterns is to find the trajectory patterns on the networks by modeling the moving objects constrained by the road architecture [187], [188]. Several recent studies have been developed to mine predesignated trajectory patterns [189]–[194], which are defined as follows.

- *Flock*: A group of objects that travel within some disk for consecutive timestamps.
- *Moving cluster*: A set of objects that move close to each other for a long time interval.
- *Convoy*: A group of objects that are density-connected to each other within a consecutive timestamps.
- *Traveling companion*: A group of objects that its members are density connected by themselves for a period and its size is larger than a threshold.
- *Gathering*: A dense and continuing group of individuals with low mobility of individuals in this group.

Gudmundsson *et al.* [189] computed four types of spatiotemporal patterns using approximation algorithms, including flock, leadership, convergence, and encounter. Later, they detected the longest duration flocks and meetings [191]. Jeung *et al.* [192] proposed a trajectory simplification (CuTS) method with the filter-refinement framework to discover conveys. The filter step applies line-simplification techniques on the trajectories. The refinement step further process candidate convoys to obtain actual convoys. Both the flock and convoy discoveries require the moving objects to stick together for consecutive timestamps. Li *et al.* [195] considered the situation that the moving objects in a cluster might diverge temporarily and recongregate at certain timestamps and proposed a novel trajectory pattern, which is called swarm. Tang *et al.* [193] proposed a smart-and-closed discovery algorithm to efficiently generate traveling companions from trajectory data in an incremental manner. Both real taxi GPS and synthetic trajectory data sets were used to evaluate the algorithm. In recent research, Zheng *et al.* [194] implemented a framework to find gatherings. It consists of three phases: snapshot clustering, crowd discovery, and gathering detection. Snapshot clustering performs density-based clustering on the object trajectories at each timestamp to obtain snapshot clusters. Crowd discovery finds all the closed crowds from snapshot clusters. For efficiency, the closed crowds are discovered by incrementally appending the snapshot clusters to the current set of crowd candidates at the next timestamp.

Trajectory clustering is useful for discovering movement patterns that help illuminate overall trends in the trajectories. Trajectory clustering techniques aim to find groups of moving object trajectories that are close to each other and have similar geometric shapes [194]. Since trajectory data are received incrementally, e.g., continuous new points reported by the GPS systems, the clustering methods (see Section III-A1a) with incremental learning are more practical to compute efficiently and effectively. Jesen *et al.* [196] maintained a clustering of moving object trajectories in 2-D Euclidean space with an incremental clustering scheme. Li *et al.* [197] proposed an incremental clustering framework, i.e., incremental trajectory clustering using microclustering and macroclustering (TCMM). In TCMM, the microclustering step clusters the trajectory segments at fine granularity, and the microclusters are updated constantly with newly received data. The macroclustering step is on demand of a user's request and takes microclusters as input to get full trajectory clusters. The performance of TCMM was tested on taxi GPS data. Most trajectory clustering algorithms take similar trajectories as a whole and group them to discover common trajectories. To find common subtrajectories, Lee *et al.* [198] proposed a partition-and-group framework, which partitions a trajectory into a set of line segments and then groups similar line segments together into a cluster. Experiments were carried out on hurricane track data and animal movement data to find out representative trajectories.

In addition to predefined patterns, much research has been carried out to mine frequent patterns [177], [199]–[202]. In [177], to discover similar trajectory patterns, nonmetric similarity functions were formalized based on the LCSS model. The LCSS is used to measure two sequences by allowing them to stretch, without rearranging the sequence of the elements but allowing some elements to be unmatched. Based on the LCSS model, Yan [199] developed a spatial pattern discovery model, which is called network-enhanced LCSS scheme (Net-LCSS), to measure consumer shopping path similarity. Patterns in trajectories are interesting if they are sharable by multiple commuters. Gidofalvi and Pedersen [200] mined long sharable patterns in traffic trajectories. In recent literature [202], Wei *et al.* presented a route inference framework based on collective knowledge (RICK) to construct the popular routes from uncertain trajectories. Given a location sequence and a time span, the RICK is able to construct the top-k routes that sequentially pass through the locations within the specified time span, by aggregating such uncertain trajectories in a mutual reinforcement way.

The data mining community has long been working on trajectory data. They have studied different kinds of patterns [190], [203]. More works on trajectory pattern mining can be seen [204].

2) Prediction of Movement: Much research has been developed to predict the future trajectories of moving objects on the road network. The MOST model [205], which is based on the concept of a motion vector, is able to represent near future developments of moving objects. However, the predictive movement is limited to a single motion function. Aggarwal and Agrawal [206] introduced a nonlinear model that uses quadratic predictive function. Tao *et al.* [207] proposed a prediction method based on recursive motion functions for objects with unknown motion patterns. Cai and Ng [208] used Chebyshev polynomials to represent and index spatiotemporal trajectories. In [209], Tao *et al.* developed Venn sampling, a novel estimation method optimized for a set of pivot queries that reflect the distribution of actual ones. These prediction methods aim to predict the individual object location.

The aforementioned research on prediction is based on predefined prediction model. Several studies derive the probability of the possible destinations based on historical trajectories and route decomposition. Destination prediction is an important task for many LBSs, such as recommending sightseeing places and targeted advertising.

a) Destination based on history: A common approach for destination prediction is to use the historical spatial trajectories. If a partial trip matches part of a popular route, the destination is predicted according to the popular route's destination. Krumm and Horvitz [210] proposed a method, called predestination, to predict the driver's destination by using Bayesian inference based on the history of a driver's destination. In [211], considering the characteristics of a road network, a trajectory was represented as a series of road segments. A novel similarity function is devised to search similar trajectories for a given query trajectory. Then, the trajectory with the highest frequency is treated as a future path of the query. In [212], Tiesyte *et al.* proposed a nearest-neighbor trajectory technique that identifies the historical trajectory that is the most similar to the current vehicle trajectory to predict the future movement of the vehicle.

b) Destination based on route decomposition: Route decomposition is another way for destination estimation. Simmons *et al.* [213] built an HMM of the routes and destinations and made predictions of the destinations and route through online observation of GPS positions. Ziebart *et al.* [214] used the Bayesian inference with a grid representation of the road network. The vehicle route preference was queried by counting the number of trajectories that are partially matched by the query trajectory and terminate at a location n_j. In [215], they used a tree structure to represent the mined movement patterns. Then, the online movement data were matched to the movement patterns by stepping down the tree. Finally, the person's destination and routes were predicted using the matching results. In [199], to recognize consumer shopping activities and purchase interest from RFID shopping trip data, they applied two dynamic Bayesian network models with different structures. In [216], they employed both the vehicle locations and their visiting orders to accurately classify vehicle trajectories. Newly arrived trajectories can be predicted by the pretrained classification model. In the literature [217], Mathew *et al.* first clustered the location histories and trained an HMM based on the clusters. For a given sequence of visits, the most probable next location was predicted by inferring on the HMM. However, if no popular route is matched, they fail to predict the destinations; this is called a data sparsity problem. To solve this problem, Xue *et al.* [218] proposed a subtrajectory synthesis (SubSyn) algorithm with an MRF model. The SubSyn algorithm first decomposes historical trajectories into subtrajectories comprising two neighboring locations and then connects the subtrajectories into "synthesized" trajectories. As long as the query trajectory matches part of any synthesized trajectory, the destination of the synthesized trajectory can be used for destination prediction. Real-world taxi trajectories were used to test the performance of SubSyn algorithm.

3) Discovery of Anomalies: The detection of trajectory anomalies has been widely discussed and studied using GPS data. Knorr *et al.* [219] detected trajectory outlier with a distance-based method. A trajectory was represented by a sequence of key features, and the distance was measured by summing the difference of the feature values. Different from [219], Lee *et al.* [220] proposed a partition-and-detect framework that can detect outlying subtrajectories. This framework first partitions a trajectory into a set of line segments and detects outlying line segments for trajectory outliers. The advantage of this framework is that it can detect outlying subtrajectories. In the study [221], they proposed a new framework named ROAM (Rule- and Motif-based Anomaly Detection in Moving Objects). A motif is a prototypical movement pattern, such as right turn, U-turn, and loop. Features are generated with the detected motifs. A rule-based classifier was developed to classify normal and abnormal trajectories. In [222], a method for detecting temporal outliers based on historical similarity trends was presented. To monitor distance-based anomaly on moving object trajectories, Bu *et al.* [223] utilized the local continuity to build local clusters on trajectory streams. For a base window B of the trajectory, if the numbers of neighbors in the left and right sliding windows are less than a predefined threshold, B is output as an anomaly. Ge *et al.* [224]

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

proposed a TOP-EYE method to compute the outlying score of each trajectory. The monitoring area is first divided into small grids. Then, each grid is partitioned into eight direction bins to summarize the directions with a direction vector. Finally, the density of each grid is computed. A trajectory outlier is defined to be the one that deviates from most of trajectories within an observed space and time in terms of direction. In [225], to discover anomalous driving patterns from taxi's GPS traces, Zhang *et al.* first grouped the taxi trajectories with the same origin–destination pairs. Then, an isolation-based anomalous trajectory (iBAT) detection method was proposed to detect anomalous taxi trajectory patterns. Experiments with large-scale taxi data showed that iBAT achieves good performance. In [226] and [227], they proposed an isolation-based online anomaly trajectory detection (iBOAT) method to detect anomalous trajectory, which can report all the anomaly records in a query trajectory. In [228], to detect "persistent outliers" and "emerging outliers", an efficient mining approach was proposed to cater for spatiotemporal traffic data. Two statistical models were proposed, which encompass the generic features of anomalous patterns.

In addition to single trajectory outlier detection, traffic flow anomaly detection is also studied based on GPS data. Traditional traffic jam detection methods are based on roadside sensors, such as induction loops or radar and monitor only a few critical points [229]. A GPS-based method, however, can theoretically monitor a complete road network. Wang *et al.* [230] used 24 days of taxi GPS trajectories in Beijing to detect traffic jams. The GPS trajectories were cleaned from sensor errors and fix apparent errors on the road network. After estimating free flow speed on each road segment, traffic jam events were automatically detected at roads based on relative low road-speed detection.

4) Discovery of Interesting Regions: Given a geospatial region, it is important to mine interesting locations on the road network. In addition to the existing works that focus on the geometric properties of the trajectories, semantic trajectories are often used to integrate the background geometric information to trajectory sample points for the road network. The trajectories consist of a set of stops and moves. In the work of [231], a framework for mining and modeling moving patterns was proposed from a semantic point of view. Different kinds of patterns can be discovered considering stops and moves in trajectories: 1) the most frequent stops during a certain period of time; 2) frequent stops that have duration higher than a given threshold; 3) frequent moves at a certain time interval; 4) most frequent moves inside a certain region; 5) frequent moves that intersect a given spatial feature type, and so on. They developed the novel and efficient framework to find one specific kind of pattern and frequent moves between two stops. A clustering-based algorithm was proposed to identify stops and moves of trajectories, which is called clustering-based SMoT (CB-SMoT) [232]. In the experiments, this method took buildings and the geographic data corresponding to areas as candidate stops and found the stops efficiently. To mine interesting locations, such as Tiananmen Square in Beijing, and classical travel sequences between these locations, Zheng *et al.* [233] first modeled multiple individuals' location histories by using a tree-based hierarchical graph (TBHG). A TBHG is the integration of two structures, i.e., a tree-based hierarchy H and a graph G on each level of this tree. The tree expresses the parent–children (or ascendant–descendant) relationship of the nodes pertaining to different levels, and the graphs specify the peer relationships among the nodes on the same level. Then, based on the TBHG architecture, they proposed a hypertext-induced-topic-search-based inference model, which regards an individual's access on a location as a directed link from the user to that location. This model infers a user's travel experience and the interest of a location. This system was evaluated with a real-world GPS data set collected from 107 users over a period of one year.

Finding hot routes (traffic flow patterns) on the road network is an important problem. They are beneficial to city planners, police departments, real estate developers, and many others. Knowing the hot routes allows the city to better direct traffic or analyze congestion causes. If vehicles traveled in organized clusters, it would be straightforward to use a clustering algorithm to find the hot routes. However, in the real world, vehicles move in unpredictable ways. Variations in speed, time, route, and other factors cause them to travel in rather fleeting "clusters". Li *et al.* [234] proposed a density-based algorithm, named FlowScan, to discover hot routes in the city. It handles the complexities in the trajectory data, and they performed extensive experiments verify the robustness of the algorithm. In this research [235], hierarchical travel experience information of road networks was given by statistical analysis on a large amount of taxi GPS trajectories. According to taxi trajectories, the roads can be divided into frequent roads, secondary frequent roads, and seldom roads. These trajectories can well reflect a hierarchical cognition of road networks by considering taxi driver's cognition of the road network.

IV. Image Acquisition and ITS Services

In this section, we will discuss the layers of image acquisition and ITS services, which are related to practical applications.

A. Image Acquisition of Traffic Scenes

In the layer of image acquisition, the characteristics of traffic scenes and imaging technologies are two important aspects for video-based traffic surveillance.

1) Characteristics of Traffic Scenes: From the perspective of traffic surveillance, we analyze the characteristics of four typical traffic scenes, including the road intersection, road section, highway, and tunnel, to help guide algorithm design for specific challenges.

- Road intersection: At road intersections, vehicles often turn right, left, and around, which lead to various vehicle poses, which cause recognition challenges. In addition to vehicle objects, there are also many other objects of interest, such as pedestrians, bicyclists, traffic signs, and other infrastructures. Robust vehicle detectors must consider variations between these different classes of intersection objects.
- Road section: In road sections (mid-block), vehicles usually travel along only in the lane directions. During heavy

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VI
CAMERA PARAMETERS OF SOME REPRESENTATIVE DEVICES DEPLOYED IN EXISTING VEHICLE SURVEILLANCE SYSTEMS

Device	**Applications**	**Camera Parameters**			**Traffic Scenarios**
		Resolution	**Frame Rate**	**Camera Coverage**	
VIRAT [239]	Person and vehicle events	1920×1080	30 FPS	-	City road, parking lot
i-LIDS [240]	Parked vehicle detection	CIF4 (576×704)	25 FPS	2 lanes	Road section
Autoscope Pn-530 [241]	Traffic data collection and Incidents detection	PAL: 752×582 NTSC: 768×494	-	-	Tunnels, Highways, Bridges
Autoscope Ex-140 [241]	License plate recognition	1920×1080	50 FPS	ANPR range: distance up to 40 meters	Parking; Travel time monitoring; Police rapid deployment
Citilog SmartCam [242]	Automatic incident detection and traffic data collection	1280×960	30 FPS	-	Bridges; freeways, tollways, ringroads
Kapsch VR-2 [243]	License plate recognition	1392×1032	-	Field-of-view 4m at distance 11.5m	Depending on the actual application
Kapsch VDR [243]	Vehicle detection and registration	1280×1024	-	Field-of-view 4m at distance 17m	Depending on the actual application
Traficon TRAFIBOT [244]	Automatic incident detection and traffic data collection	D1 (720×480)	30 FPS	-	Tunnel, Highway, Bridge
Traficon VIP-TX [244]	Automatic incident detection and traffic data collection	VGA (640×480)	30 FPS	-	-
Traficon EYE-D [244]	Automatic number plate recognition	1280×1024	-	2 lanes up to 60 meter distance	Automatic tolling, urban congestion charging
HIKVISION [245]	Illegal behavior detecting and recording; ANPR	2560×2048	25 FPS	3 lanes	Road intersection
DAHUA [246]	Illegal behavior detecting and recording	2432×2048	25 FPS	3 lanes	Road intersection
Iteris SmartSpan [247]	Stop-bar detection and Traffic data collection	768×494	-	Lens 4.6-46 degrees wide	Road intersection

commute hours, vehicles move slowly or even stop, leading to long queues. In these situations, vehicle occlusion occurs frequently, and stopped vehicles are lost completely using motion-based detection techniques.

- Highway: Vehicles move fast on highways (urban expressways) and have limited time in the camera FOV. Cameras need to have a high frame rate, and detection, tracking, and analysis algorithms must be efficient for real-time performance.
- City tunnel: City tunnel scenarios are similar to nighttime conditions with some supplemental lighting. However, the lighting is localized and does not provide uniform coverage as under sunlight.

2) Imaging Technology for Image Acquisition: As the basis of vision-based surveillance systems, the task of this layer is to obtain images from traffic scenes using image sensors. Because of improvements in image sensor technology in recent years, captured traffic images now have higher resolution and better image quality. Clearer images make it possible to recognize more detailed information about vehicles and other objects.

In the context of traffic surveillance, it is crucial to investigate the devices that are currently deployed in existing vehicle surveillance systems. In Table VI, we list the camera parameters of some representative devices with respect to applications and traffic scenarios, including camera resolution, frame rate [frames per second (FPS)], and camera coverage.

With the great progress of image sensors, more advanced cameras may be deployed for traffic surveillance in the next decade. For example, Sony has developed a 20.68 mega-pixel (5256 × 3934) high-speed CMOS image sensor [245] with the frame rate of 22 FPS. ON Semiconductor VITA 25 K series image sensor [246] can achieve 53 FPS at full resolution of 5120 × 5120 pixel. With higher resolution images, we can obtain more detailed information about all visible vehicles, such as vehicle license plate numbers, model badging, or parking stickers. A high frame rate enables capture of faster vehicles and better estimation of slight vehicle movements such as braking or rapid starting.

B. ITS Services

In this section, we present a few ITS services that will highlight the impact of video-based networked vehicle surveillance system, including security monitoring, traffic flow analysis, and environment impact assessment.

- Illegal activity and anomaly detection: Illegal driving leads to traffic accidents and causes hidden traffic troubles. By means of video-based surveillance, illegal driving behaviors, i.e., red light running and reverse driving, will be detected, and the evidence will be reported to authorities.
- Security monitoring: To monitor a specific vehicle of interest, the network-based surveillance system can report the trajectory on the road network. This functionality along with real video streams can help law enforcement monitor activities and prevent crimes.
- Electronic toll collection: When a vehicle travels through charging ports, i.e., the entrances and exits of the highway or the parking port, a camera sensor will recognize the attributes of the vehicle, particularly its license plate number. A toll system can then be implemented based on

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

visual recognition results rather than currently more popular RFID-based toll methods. Electronic toll collection becomes more efficient and reliable when using vision systems since the toll is assessed to a vehicle and not the RFID tag, which can be compromised by fraudulent activities.

- Traffic flow analysis: Real-time traffic information should be provided to drivers in a timely manner to better manage congestion due to incidents such as emergency events, special events, weather, work zones, or daily commute patterns. By analyzing traffic flow, ITS systems can find congested road sections and regions; furthermore, the traffic flow might be predicted, and road users can be guided to alternative routes before traffic jams.
- Transportation planning and road construction: A networked surveillance system can report the traffic pattern and find the traffic jams and other anomaly of traffic flow. This information might be fed back to the department of transportation to provide suggestive data transportation planning and road construction.
- Environment impact assessment: Different types of vehicles, e.g., truck, bus, and car, will bring different influences to the environments. Impact models of environment influences for all types of vehicles can be constructed, respectively. Environment impact assessment can be done by analyzing and forecasting the trend of the traffic flow using the networked video surveillance system.

V. Discussion of Future Developments

In the previous sections, we have provided the state-of-the-art techniques for vehicle surveillance in ITS. There are many open issues, and further research will be carried out to fully realize the promise of video traffic surveillance systems. Here, we provide our perspective on future developments of networked video surveillance systems.

1) Improving the System Performance: The performance of most existing surveillance systems will degrade in complex traffic scenarios (e.g., vehicle occlusion, pose variation, and illumination change), as listed in Section I. In the previous sections, we have discussed the challenges with respect to each module of video surveillance system and the corresponding existing solutions. Here, we describe our views of some special challenging issues.

a) Occlusion handling: Vehicle occlusion is caused by the mapping from 3-D traffic scenes into 2-D images, which result in loss of vehicle's visual information [58]. This will lead to false detection of the occluded objects. Common methods of occlusion handling are to use the visual information of the object to detect it while ignoring the features of the occluded parts [43], [53], [58], [60], [247]. Overall, vehicle occlusion can be handled with two steps. First, the system should determine the presence of occlusion. Often times, object occlusion is a gradual process, and one can infer the presence of occlusion by observing previous detection results. The presence of occlusion can also be determined by the quality of response of the object detection model [248], [249] in a still image. Second, the occluded object can be detected by explicit occlusion handling. One common practice is the use machine learning methods to learn the model of the occluded object samples [250] and detect with the learned model. The other method is to learn the object model without occlusion and detect with a designated mask.

b) Dealing with pose variation and different vehicle types: Vehicles often change their poses while traveling on the road (e.g., changing lanes and turn). This results in completely different appearance in the image for the same object. The dual problem that arises during video surveillance is the apparent appearance similarity between two different vehicle types, such as between a van and SUV. The wide variability in intravehicle appearance with little intervehicle differentiation makes appearance-based algorithms difficult to apply in practice. For example, detectors can be trained with different models for different poses, but this increases complexity and affects real-time performance. Although, motion-based detection will be not affected by these influence, the detected moving object might be not a vehicle, and it is susceptible to the shadow problem. In these cases, shadows must be explicitly detected and removed for implementation [54]–[57].

c) Adapting to illumination change: Various weather patterns and different times of day lead to illumination variation and yield great differences in object appearance. Under strong lighting conditions, the texture of the vehicle is obvious, while much information of the vehicle is not visible under insufficient lighting conditions, e.g., at nighttime. In order to suppress the influence of illumination change, features that are not sensitive to the lighting condition are usually used, including SIFT [21] and HOG [23]. Particularly during the nighttime, most features of the vehicle are not visible. One solution is to use additional supplemental lighting equipment for the camera or focus on the only visible parts of the vehicle, such as the headlight and the taillight [61]–[65].

d) Surveillance with CVISs.: With the development of telematics and cooperative vehicle–infrastructure systems (CVISs), the communication between a vehicle and roadside infrastructure, including cameras, becomes feasible. Usually, CVIS is used to publish warnings and other information to vehicles. In our view, the roadside camera can detect the vehicles by combining with the sensors on the vehicle, such as the vehicle cameras. For example, in the case of vehicle occlusion, the cameras on the nonoccluded vehicles can be used to detect the occluded vehicles and transmit the detection result to the roadside camera [251]. By this method, the surveillance system is strengthened. Similar approaches are worth studying in future research.

2) Networked Surveillance System: Single-camera-based surveillance systems can only monitor traffic objects in the FOV of the camera, limiting global awareness. With the great advances of network technology and the Internet of Things, there is a trend that the cameras on the road are networked. The HNVS framework presented in this paper not only monitors a vehicle's behavior at a single camera node but also analyzes its behavior over the road network. Since the physical location of cameras on the road network is fixed, they can serve as an LBS. Currently, RFID and GPS are the most commonly used sensors to obtain the object trajectories within a wide range. (GPS is the primary means for extracting vehicle trajectories.) Researchers have carried out many GPS-based trajectory analyses, as shown in Section III-B; however, these studies are mostly based on fleet vehicles, such as taxis or trucks, which may not reflect "typical" driving patterns. Compared with the GPS sensors, which mainly collect the location information, the networked surveillance system has the following characteristics.

- Similar to a GPS sensor, a camera can report the vehicle location on the road network. The networked system can only obtain the discrete locations within deployed camera views, whereas GPS can obtain continuous trajectory. However, vehicle behaviors on the road network are mostly analyzed based on specific road sections, so that the granularity of the camera network is sufficient for network behavior analysis.
- The networked surveillance system can also perceive the detailed characteristics of the vehicle, including license plate number, vehicle color, vehicle type, vehicle logo, etc., as shown in Section II-C. Thus, the trajectory analysis can be performed according to these attributes, e.g., analyzing the bus trajectories, the car trajectories, and even analyzing the vehicles with various manufacturers and colors. From this perspective, the networked system is more flexible than the GPS-based system, making the study of trajectory pattern discovery, motion prediction, anomaly detection, and other issues worthwhile.

While there have been studies on multicamera tracking, they typically relay on overlapped or spatially adjacent cameras that cannot always be used on road networks since cameras are often far in distance. Vehicle reidentification techniques to track the same vehicle over these large distances are applicable in these scenarios [252]. The camera usually needs to maintain the models of the objects detected by other cameras to perform reidentification task. For the networked cameras on the road network, the topology of the networked cameras is difficult to obtain, and the number of camera nodes is large, making maintenance of object models across all cameras difficult. One solution is to perform multicamera tracking within the neighborhood of each camera because a vehicle should usually drive from a camera node to one of its neighborhood cameras. Another more convenient way is for the cameras to transmit the detection results to a remote central server, and the trajectory is obtained based on the vehicles' dynamic and static attributes by the server.

3) Deeper Understanding of Traffic Scenes: Behavior understanding is defined as the analysis and interpretation of individual behaviors and interactions between objects for visual surveillance [154]. From our point of view, traffic surveillance is to monitor the dynamic and static attributes of the traffic objects and then analyze how they affect the traffic scenes in feedback. It is feasible to perform deeper understanding of traffic scenes with a networked surveillance system.

1) The dynamic and static attributes of each vehicle driving on the road should be extracted and analyzed, including their attributes on the road network, as described in Sections II and III.
2) The complete transportation and traffic situation should be considered. For traffic surveillance, other traffic objects such as pedestrians, traffic signals, and traffic signs can be detected to further understand the vehicle behavior. For example, a red light run behavior can be recognized by considering both the vehicle trajectory and the state of the traffic signal.
3) Higher level transportation analysis can be built on top of basic vehicle monitoring. As an example, transportation environmental pollution impacts can be assessed along the road network by utilizing emission models for specific vehicles along with vehicle tracking and type classification [253]. In addition, regular patterns of traffic flow and traffic jams can be discovered with the vehicle trajectory analysis. This can provide suggestions about traffic flow induction and road design in turn.

VI. Conclusion

The video-based traffic surveillance system has become an important part in ITSs. These systems can acquire the images of traffic scenes, analyze the information of the traffic objects, and understand their behaviors and activities. In this paper, we have presented the HNVS architecture in ITSs to review the state-of-the-art literature. The aim of vehicle surveillance is to extract the vehicles' attributes and understand vehicles' behaviors. Based on this consideration, the HNVS framework first extracts vehicles' dynamic and static attributes by a single camera node and then analyzes vehicles' behaviors on the road network. HNVS is both hierarchical and networked. First, the functions have very little overlap between different layers. Second, HNVS is a networked surveillance framework that makes it possible to capture and understand the vehicle behaviors over the entire road network. We have provided a comprehensive survey of the state-of-the-art methods on attribute extraction, including techniques of vehicle detection, tracking, recognition, and tracking on the road network. Detailed discussions on the challenges are carried out with each module. Then, research related to vehicle behavior understanding is reviewed, and it is noted that most research about network vehicle behavior understanding is not based on video sensors, but there is potential for application in networked vehicle surveillance systems. Further achievements in this field of research will provide more effective ITS services for widespread implementation.

Acknowledgment

We would like to express our sincere appreciation to professor Fei-Yue Wang for his instruction and encouragement.

References

[1] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920–939, Sep. 2011.
[2] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, Jan. 2013.
[3] L.-W. Tsai, J.-W. Hsieh, and K.-C. Fan, "Vehicle detection using normalized color and edge map," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 850–864, Mar. 2007.

[4] R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 119–130, Jun. 2000.

[5] S. Gupte, O. Masoud, R. F. K. Martin, and N. P. Papanikolopoulos, "Detection and classification of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 1, pp. 37–47, Mar. 2002.

[6] A. Ottlik and H.-H. Nagel, "Initialization of model-based vehicle tracking in video sequences of inner-city intersections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 211–225, Nov. 2008.

[7] X. Ma and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1185–1192.

[8] N. Buch, J. Orwell, and S. Velastin, "3D extended Histogram of Oriented Gradients (3DHOG) for classification of road users in urban scenes," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2009, pp. 1–11.

[9] R. S. Feris *et al.*, "Large-scale vehicle detection, indexing, and search in urban surveillance videos," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 28–42, Feb. 2012.

[10] G. Marola, "Using symmetry for detecting and locating objects in a picture," *Comput. Vis. Graph. Image Process.*, vol. 46, no. 2, pp. 179–195, May 1989.

[11] A. Kuehnle, "Symmetry-based recognition of vehicle rears," *Pattern Recognit. Lett.*, vol. 12, no. 4, pp. 249–258, Apr. 1991.

[12] W. Von Seelen, C. Curio, J. Gayko, U. Handmann, and T. Kalinke, "Scene analysis and organization of behavior in driver assistance systems," in *Proc. Int. Conf. Image Process.*, 2000, vol. 3, pp. 524–527.

[13] D. Guo, T. Fraichard, M. Xie, and C. Laugier, "Color modeling by spherical influence field in sensing driving environment," in *Proc. IEEE Intell. Veh. Symp.*, 2000, pp. 249–254.

[14] N. Matthews, P. An, D. Charnley, and C. Harris, "Vehicle detection and recognition in greyscale imagery," *Control Eng. Pract.*, vol. 4, no. 4, pp. 473–479, Apr. 1996.

[15] C. Goerick, D. Noll, and M. Werner, "Artificial neural networks in real-time car detection and tracking applications," *Pattern Recognit. Lett.*, vol. 17, no. 4, pp. 335–343, Apr. 1996.

[16] N. Srinivasa, "Vision-based vehicle detection and tracking method for forward collision warning in automobiles," in *Proc. IEEE Intell. Veh. Symp.*, 2002, vol. 2, pp. 626–631.

[17] M. Bertozzi, A. Broggi, and S. Castelluccio, "A real-time oriented system for vehicle detection," *J. Syst. Archit.*, vol. 43, no. 1–5, pp. 317–325, Mar. 1997.

[18] T. Kalinke, C. Tzomakas, and W. V. Seelen, "A texture-based object detection and an adaptive model-based classification," in *Proc. IEEE Int. Conf. Intell. Veh.*, 1998, pp. 143–148.

[19] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[20] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1475–1490, Nov. 2004.

[21] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

[22] Z. Kim and J. Malik, "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 524–531.

[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 1, pp. 886–893.

[24] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. 511–518.

[25] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[26] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2553–2561.

[27] R. Socher, B. Huval, B. P. Bath, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3D object classification," in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 665–673.

[28] Z. Chen, T. Ellis, and S. A. Velastin, "Vehicle detection, tracking and classification in urban traffic," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2012, pp. 951–956.

[29] Y. Freund, "Boosting a weak learning algorithm by majority," *Inf. Comput.*, vol. 121, no. 2, pp. 256–285, Sep. 1995.

[30] J. Winn and J. Shotton, "The layout consistent random field for recognizing and segmenting partially occluded objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2006, vol. 1, pp. 37–44.

[31] D. Hoiem, C. Rother, and J. Winn, "3D layoutcrf for multi-view object class recognition and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.

[32] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 108–118, Jun. 2000.

[33] C.-C. R. Wang and J. J. J. Lien, "Automatic vehicle detection using local features: A statistical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 83–96, Mar. 2008.

[34] Y. N. Wu, Z. Si, H. Gong, and S.-C. Zhu, "Learning active basis model for object detection and recognition," *Int. J. Comput. Vis.*, vol. 90, no. 2, pp. 198–235, Nov. 2010.

[35] S. Sivaraman and M. M. Trivedi, "Vehicle detection by independent parts for urban driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1597–1608, Dec. 2013.

[36] L. Lin, T. Wu, J. Porway, and Z. Xu, "A stochastic graph grammar for compositional object representation and recognition," *Pattern Recognit.*, vol. 42, no. 7, pp. 1297–1307, Jul. 2009.

[37] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[38] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 2241–2248.

[39] H. T. Niknejad, T. Kawano, Y. Oishi, and S. Mita, "Occlusion handling using discriminative model of trained part templates and conditional random field," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 750–755.

[40] L. Yang, B. Yao, W. Yongtian, and S.-C. Zhu, "Reconfigurable templates for robust vehicle detection and classification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2012, pp. 321–328.

[41] G. D. Sullivan, K. D. Baker, A. D. Worrall, C. Attwood, and P. Remagnino, "Model-based vehicle detection and classification using orthographic approximations," *Image Vis. Comput.*, vol. 15, no. 8, pp. 649–654, Aug. 1997.

[42] S. Messelodi, C. M. Modena, and M. Zanin, "A computer vision system for the detection and classification of vehicles at urban road intersections," *Pattern Anal. Appl.*, vol. 8, no. 1/2, pp. 17–31, Sep. 2005.

[43] X. Song and R. Nevatia, "Detection and tracking of moving vehicles in crowded scenes," in *Proc. IEEE Workshop Motion Video Comput.*, 2007, pp. 1–4.

[44] B. Johansson, J. Wiklund, P.-E. Forssén, and G. Granlund, "Combining shadow detection and simulation for estimation of vehicle size and position," *Pattern Recognit. Lett.*, vol. 30, no. 8, pp. 751–759, Jun. 2009.

[45] M. J. Leotta and J. L. Mundy, "Predicting high resolution image edges with a generic, adaptive, 3-D vehicle model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 1311–1318.

[46] J. Liebelt, C. Schmid, and K. Schertler, "Viewpoint-independent object class detection using 3D feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[47] A. Lai and N. H. C. Yung, "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1998, vol. 4, pp. 241–244.

[48] P. Kumar, S. Ranganath, and H. Weimin, "Bayesian network based computer vision algorithm for traffic monitoring using video," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2003, vol. 1, pp. 897–902.

[49] B. T. Morris and M. M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, Sep. 2008.

[50] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1999, vol. 2, p. 252.

[51] J. Zheng, Y. Wang, N. L. Nihan, and M. E. Hallenbeck, "Extracting roadway background image: Mode-based approach," *J. Transp. Res. Board*, vol. 1944, no. 1, pp. 82–88, 2006.

[52] T. Gao, Z.-G. Liu, W.-C. Gao, and J. Zhang, "A robust technique for background subtraction in traffic video," in *Advances in Neuro-Information Processing*. Berlin, Germany: Springer-Verlag, 2009, pp. 736–744.

[53] C.-L. Huang and W.-C. Liao, "A vision-based vehicle identification system," in *Proc. Int. Conf. Pattern Recog.*, 2004, vol. 4, pp. 364–367.

[54] A. Yoneyama, C.-H. Yeh, and C.-C. J. Kuo, "Robust vehicle and traffic information extraction for highway surveillance," *EURASIP J. Appl. Signal Process.*, vol. 14, no. 1, pp. 2305–2321, Jan. 2005.

[55] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[56] X. Ji, Z. Wei, and Y. Feng, "Effective vehicle detection technique for traffic surveillance systems," *J. Vis. Commun. Image Represent.*, vol. 17, no. 3, pp. 647–658, Jun. 2006.

[57] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 918–923, Jul. 2003.

[58] W. Zhang, Q. J. Wu, X. Yang, and X. Fang, "Multilevel framework to detect and handle vehicle occlusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 161–174, Mar. 2008.

[59] N. K. Kanhere and S. T. Birchfield, "Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 148–160, Mar. 2008.

[60] C. C. C. Pang, W. W. L. Lam, and N. H. C. Yung, "A method for vehicle count in the presence of multiple-vehicle occlusions in traffic images," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 441–459, Sep. 2007.

[61] G. Gritsch, N. Donath, B. Kohn, and M. Litzenberger, "Night-time vehicle classification with an embedded, vision system," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2009, pp. 1–6.

[62] K. Robert, "Video-based traffic monitoring at day and night vehicle features detection tracking," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2009, pp. 1–6.

[63] K. Robert, "Night-time traffic surveillance: A robust framework for multi-vehicle detection, classification and tracking," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 1–6.

[64] W. Wang, C. Shen, J. Zhang, and S. Paisitkriangkrai, "A two-layer night-time vehicle detector," in *Proc. Digit. Image Comput., Tech. Appl.*, 2009, pp. 162–167.

[65] Y.-L. Chen, B.-F. Wu, and C.-J. Fan, "Real-time vision-based multiple vehicle detection and tracking for nighttime traffic surveillance," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2009, pp. 3352–3358.

[66] B.-F. Wu, Y.-L. Chen, and C.-C. Chiu, "A discriminant analysis based recursive automatic thresholding approach for image segmentation," *IEICE Trans. Inf. Syst.*, vol. E88-D, no. 7, pp. 1716–1723, Jul. 2005.

[67] W. Zhang, Q. M. J. Wu, G. Wang, and X. You, "Tracking and pairing vehicle headlight in night scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 140–153, Mar. 2012.

[68] L. Unzueta *et al.*, "Adaptive multicue background subtraction for robust vehicle counting and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 527–540, Jun. 2012.

[69] F. Bardet and T. Chateau, "MCMC particle filter for real-time visual tracking of vehicles," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2008, pp. 539–544.

[70] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 19–23.

[71] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 809–817.

[72] D. Roller, K. Daniilidis, and H.-H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *Int. J. Comput. Vis.*, vol. 10, no. 3, pp. 257–281, Jun. 1993.

[73] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[74] R. Rad and M. Jamzad, "Real time classification and tracking of multiple vehicles in highways," *Pattern Recognit. Lett.*, vol. 26, no. 10, pp. 1597–1607, Jul. 2005.

[75] J. Lou, H. Yang, W. M. Hu, and T. Tan, "Visual vehicle tracking using an improved EKF," in *Proc. Asian Conf. Comput. Vis.*, 2002, pp. 296–301.

[76] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[77] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vis. Comput.*, vol. 21, no. 1, pp. 99–110, Jan. 2003.

[78] T. Mauthner, M. Donoser, and H. Bischof, "Robust tracking of spatial related components," in *Proc. Int. Conf. Pattern Recog.*, 2008, pp. 1–4.

[79] S. Khatoonabadi and I. Bajic, "Video object tracking in the compressed domain using spatio-temporal Markov random fields," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 300–313, Jan. 2013.

[80] D. Zheng, Y. Zhao, and J. Wang, "An efficient method of license plate location," *Pattern Recognit. Lett.*, vol. 26, no. 15, pp. 2431–2438, Nov. 2005.

[81] S.-Z. Wang and H.-J. Lee, "Detection and recognition of license plate characters with different appearances," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2003, vol. 2, pp. 979–984.

[82] F. Faradji, A. H. Rezaie, and M. Ziaratban, "A morphological-based license plate location," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 1, pp. 57–60.

[83] C. N. K. Babu and K. Nallaperumal, "An efficient geometric feature based license plate localization and recognition," *Int. J. Imag. Sci. Eng.*, vol. 2, no. 2, pp. 189–194, Apr. 2008.

[84] V. Kamat and S. Ganesan, "An efficient implementation of the hough transform for detecting vehicle license plates using DSP's," in *Proc. Real-Time Technol. Appl. Symp.*, 1995, pp. 58–59.

[85] A. M. Al-Ghaili, S. Mashohor, A. Ismail, and A. R. Ramli, "A new vertical edge detection algorithm and its application," in *Proc. Int. Conf. Comput. Eng. Syst.*, 2008, pp. 204–209.

[86] H. Zhang, W. Jia, X. He, and Q. Wu, "Learning-based license plate detection using global and local features," in *Proc. Int. Conf. Pattern Recog.*, 2006, vol. 2, pp. 1102–1105.

[87] W. Le and S. Li, "A hybrid license plate extraction method for complex scenes," in *Proc. Int. Conf. Pattern Recog.*, 2006, vol. 2, pp. 324–327.

[88] L. Dlagnekov, "License plate detection using adaboost," Dept. Comput. Sci. Eng., Univ. California San Diego, La Jolla, CA, USA, 2004.

[89] J. Jiao, Q. Ye, and Q. Huang, "A configurable method for multi-style license plate recognition," *Pattern Recognit.*, vol. 42, no. 3, pp. 358–369, Mar. 2009.

[90] X. Shi, W. Zhao, and Y. Shen, "Automatic license plate recognition system based on color image processing," in *Computational Science and Its Applications*. Berlin, Germany: Springer-Verlag, 2005, pp. 1159–1168.

[91] G. Cao, J. Chen, and J. Jiang, "An adaptive approach to vehicle license plate localization," in *Proc. IEEE Annu. Conf. Ind. Electron. Soc.*, 2003, vol. 2, pp. 1786–1791.

[92] X. Wan, J. Liu, and J. Liu, "A vehicle license plate localization method using color barycenters hexagon model," in *Proc. SPIE Int. Soc. Opt. Photon.*, 2011, pp. 80092O-1–80092O-5.

[93] K. Deb and K.-H. Jo, "A vehicle license plate detection method for intelligent transportation system applications," *Cybern. Syst.*, vol. 40, no. 8, pp. 689–705, Nov. 2009.

[94] D. Llorens, A. Marzal, V. Palazon, and J. M. Vilar, "Car license plates extraction and recognition based on connected components analysis and hmm decoding," in *Pattern Recognition and Image Analysis*. Berlin, Germany: Springer-Verlag, 2005, pp. 571–578.

[95] K. K. Kim, K. Kim, J. Kim, and H. Kim, "Learning-based approach for license plate recognition," in *Proc. IEEE Signal Process. Soc. Workshop Neur. Netw. Signal Process.*, 2000, vol. 2, pp. 614–623.

[96] M.-P. D. Jolly, S. Lakshmanan, and A. K. Jain, "Vehicle segmentation and classification using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 293–308, Mar. 1996.

[97] D. T. Munroe and M. G. Madden, "Multi-class and single-class classification approaches to vehicle model recognition from images," in *Proc. Irish Conf. Artif. Intell. Cogn. Sci.*, Londonderry, U.K., 2005.

[98] I. Zafar, E. A. Edirisinghe, S. Acar, and H. E. Bez, "Two-dimensional statistical linear discriminant analysis for real-time robust vehicle-type recognition," in *Proc. SPIE Real-Time Image Process.*, 2007, pp. 649 602-1–649 602-8.

[99] H. Huang, Q. Zhao, Y. Jia, and S. Tang, "A 2dlda based algorithm for real time vehicle type recognition," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2008, pp. 298–303.

[100] P. Negri, X. Clady, M. Milgram, and R. Poulenard, "An oriented-contour point based voting algorithm for vehicle type classification," in *Proc. Int. Conf. Pattern Recog.*, 2006, vol. 1, pp. 574–577.

[101] R. Feris, J. Petterson, B. Siddiquie, L. Brown, and S. Pankanti, "Large-scale vehicle detection in challenging urban surveillance environments," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2011, pp. 527–533.

[102] B. Zhang and C. Zhao, "Classification of vehicle make by combined features and random subspace ensemble," in *Proc. Int. Conf. Image Graph.*, 2011, pp. 920–925.

[103] O. Hasegawa and T. Kanade, "Type classification, color estimation, and specific target detection of moving targets on public streets," *Mach. Vis. Appl.*, vol. 16, no. 2, pp. 116–121, Feb. 2005.

[104] N. Baek, S.-M. Park, K.-J. Kim, and S.-B. Park, "Vehicle color classification based on the support vector machine method," in *Communications in Computer and Information Science*, vol. 2. Berlin, Germany: Springer-Verlag, 2007, pp. 1133–1139.

[105] E. Dule, M. Gokmen, and M. S. Beratoglu, "A convenient feature vector construction for vehicle color recognition," in *Proc. Int. Conf. Neural Netw. WSEAS*, 2010, pp. 250–255.

[106] H. J. Lee, "Neural network approach to identify model of vehicles," in *Advances in Neural Networks*. Berlin, Germany: Springer-Verlag, 2006, pp. 66–72.
[107] A. Psyllos, C. Anagnostopoulos, E. Kayafas, and V. Loumos, "Image processing & artificial neural networks for vehicle make and model recognition," in *Proc. Conf. Appl. Adv. Tech. Transp.*, 2008, vol. 5, pp. 4229–4243.
[108] A. Psyllos, C. Anagnostopoulos, and E. Kayafas, "Vehicle authentication from digital image measurements," in *Proc. 16th IMEKO TC4 Symp.*, 2008, pp. 1–6.
[109] Y. Wang, N. Li, and Y. Wu, "Application of edge testing operator in vehicle logo recognition," in *Proc. Int. Workshop Intell. Syst. Appl.*, 2010, pp. 1–3.
[110] Y. Wang, Z. Liu, and F. Xiao, "A fast coarse-to-fine vehicle logo detection and recognition method," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2007, pp. 691–696.
[111] H.-K. Xu, F.-H. Yu, J.-H. Jiao, and H.-S. Song, "A new approach of the vehicle license plate location," in *Proc. Int. Conf. Parallel Distrib. Comput. Appl. Technol.*, 2005, pp. 1055–1057.
[112] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 377–392, Sep. 2006.
[113] S. Yohimori, Y. Mitsukura, M. Fukumi, N. Akamatsu, and N. Pedrycz, "License plate detection system by using threshold function and improved template matching method," in *Proc. IEEE Annu. Meet. Fuzzy Inf.*, 2004, vol. 1, pp. 357–362.
[114] W. Jia, H. Zhang, X. He, and Q. Wu, "Gaussian weighted histogram intersection for license plate classification," in *Proc. Int. Conf. Pattern Recog.*, 2006, vol. 3, pp. 574–577.
[115] J.-F. Xu, S.-F. Li, and M.-S. Yu, "Car license plate extraction using color and edge information," in *Proc. IEEE Int. Symp. Ind. Electron.*, 2004, vol. 6, pp. 3904–3907.
[116] Z.-X. Chen, C.-Y. Liu, F.-L. Chang, and G.-Y. Wang, "Automatic license-plate location and recognition based on feature salience," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3781–3785, Sep. 2009.
[117] Y. Lee *et al.*, "License plate detection using local structure patterns," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2010, pp. 574–579.
[118] Q. Gao, X. Wang, and G. Xie, "License plate recognition based on prior knowledge," in *Proc. IEEE Int. Conf. Autom. Logist.*, 2007, pp. 2964–2968.
[119] J.-M. Guo and Y.-F. Liu, "License plate localization and character segmentation with feedback self-learning and hybrid binarization techniques," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1417–1424, May 2008.
[120] T. Nukano, M. Fukumi, and M. Khalid, "Vehicle license plate character recognition by neural networks," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, 2004, pp. 771–775.
[121] V. Shapiro and G. Gluhchev, "Multinational license plate recognition system: Segmentation and classification," in *Proc. Int. Conf. Pattern Recog.*, 2004, vol. 4, pp. 352–355.
[122] B.-F. Wu, S.-P. Lin, and C.-C. Chiu, "Extracting characters from real vehicle licence plates out-of-doors," *IET Comput. Vis.*, vol. 1, no. 1, pp. 2–10, Mar. 2007.
[123] P. Soille, *Morphological Image Analysis: Principles and Applications*. New York, NY, USA: Springer-Verlag, 2003.
[124] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, and T. Shiose, "A novel adaptive morphological approach for degraded character image segmentation," *Pattern Recognit.*, vol. 38, no. 11, pp. 1961–1975, Nov. 2005.
[125] M.-S. Pan, J.-B. Yan, and Z.-H. Xiao, "Vehicle license plate character segmentation," *Int. J. Autom. Comput.*, vol. 5, no. 4, pp. 425–432, Oct. 2008.
[126] Z. Ping and C. Lihui, "A novel feature extraction method and hybrid tree classification for handwritten numeral recognition," *Pattern Recognit. Lett.*, vol. 23, no. 1–3, pp. 45–56, Jan. 2002.
[127] H. Erdinc Kocer and K. Kursat Cevik, "Artificial neural networks based vehicle license plate recognition," *Proc. Comput. Sci.*, vol. 3, pp. 1033–1037, 2011.
[128] J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," *Pattern Recognit.*, vol. 28, no. 2, pp. 153–160, Feb. 1995.
[129] Y. S. Huang and C. Y. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 90–94, Jan. 1995.
[130] H.-J. Kang, K. Kim, and J. H. Kim, "A framework for probabilistic combination of multiple classifiers at an abstract level," *Eng. Appl. Artif. Intell.*, vol. 10, no. 4, pp. 379–385, Aug. 1997.
[131] C.-N. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 377–391, Sep. 2008.
[132] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.
[133] B. Morris, C. Tran, G. Scora, M. Trivedi, and M. Barth, "Real-time video-based traffic measurement and visualization system for energy/emissions," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1667–1678, Dec. 2012.
[134] R. P. Avery, Y. Wang, and G. S. Rutherford, "Length-based vehicle classification using images from uncalibrated video cameras," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2004, pp. 737–742.
[135] D. Han, M. J. Leotta, D. B. Cooper, and J. L. Mundy, "Vehicle class recognition from video-based on 3d curve probes," in *Proc. IEEE Int. Workshop Vis. Surveillance Perform. Eval. Tracking Surveillance*, 2005, pp. 285–292.
[136] V. S. Petrovic and T. F. Cootes, "Analysis of features for rigid structure vehicle type recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2004, pp. 1–10.
[137] B. Zhang, "Reliable classification of vehicle types based on cascade classifier ensembles," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 322–332, Mar. 2013.
[138] H. Yang *et al.*, "An efficient method for vehicle model identification via logo recognition," in *Proc. Int. Conf. Comput. Inf. Sci.*, 2013, pp. 1080–1083.
[139] A. P. Psyllos, C.-N. Anagnostopoulos, and E. Kayafas, "Vehicle logo recognition using a sift-based enhanced matching scheme," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 322–328, Jun. 2010.
[140] A. Minich and C. Li, "Vehicle logo recognition and classification: Feature descriptors vs. shape descriptors," Stanford Univ., Stanford, CA, USA, EE368 Final Proj., 2011.
[141] R. Eshel and Y. Moses, "Homography based multiple camera detection and tracking of people in a dense crowd," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.
[142] A. D. Straw, K. Branson, T. R. Neumann, and M. H. Dickinson, "Multi-camera real-time three-dimensional tracking of multiple flying animals," *J. R. Soc. Interface*, vol. 8, no. 56, pp. 395–409, Mar. 2011.
[143] B. C. Matei, H. S. Sawhney, and S. Samarasekera, "Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 3465–3472.
[144] R. Hamid *et al.*, "Player localization using multiple static cameras for sports visualization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 731–738.
[145] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vis. Conf.*, 2010, vol. 2, p. 6.
[146] A. Civilis, C. S. Jensen, J. Nenortaite, and S. Pakalnis, "Efficient tracking of moving objects with precision guarantees," in *Proc. Int. Conf. Mobile Ubiquitous Syst.*, 2004, pp. 164–173.
[147] A. Civilis, C. S. Jensen, and S. Pakalnis, "Techniques for efficient road-network-based tracking of moving objects," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 5, pp. 698–712, May 2005.
[148] R. Lange, F. Dü, and K. Rothermel, "Efficient real-time trajectory tracking," *VLDB J.*, vol. 20, no. 5, pp. 671–694, Oct. 2011.
[149] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.
[150] P. Borges, N. Conci, and A. Cavallaro, "Video-based human behavior understanding: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1993–2008, Nov. 2013.
[151] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *Vis. Comput.*, vol. 29, no. 10, pp. 983–1009, Oct. 2013.
[152] T. Xiang and S. Gong, "Video behavior profiling for anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 893–908, May 2008.
[153] V. S. P. Turaga, R. Chellappa, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, Nov. 2008.

[154] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.

[155] H. Veeraraghavan and N. P. Papanikolopoulos, "Learning to recognize video-based spatiotemporal events," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 628–638, Dec. 2009.

[156] W. Hu, X. Li, G. Tian, S. J. Maybank, and Z. Zhang, "An incremental dpmm-based method for trajectory clustering, modeling, and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1051–1065, May 2013.

[157] M. M. Trivedi, T. L. Gandhi, and K. S. Huang, "Distributed interactive video arrays for event capture and enhanced situational awareness," *IEEE Intell. Syst.*, vol. 20, no. 5, pp. 58–66, Sep. 2005.

[158] P. Remagnino, S. A. Velastin, G. L. Foresti, and M. Trivedi, "Novel concepts and challenges for the next generation of video surveillance systems," *Mach. Vis. Appl.*, vol. 18, no. 3, pp. 135–137, Aug. 2007.

[159] K. A. Eccles, "Automated enforcement for speeding and red light running," *J. Transp. Res. Board*, vol. 729, 2012.

[160] B. E. Porter and K. J. England, "Predicting red-light running behavior: A traffic safety study in three urban settings," *J. Safety Res.*, vol. 31, no. 1, pp. 1–8, 2000.

[161] D. Kasper *et al.*, "Object-oriented Bayesian networks for detection of lane change maneuvers," in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 673–678.

[162] S. Kamijo, H. Koo, X. Liu, K. Fujihira, and M. Sakauchi, "Development and evaluation of real-time video surveillance system on highway based on semantic hierarchy and decision surface," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2005, vol. 1, pp. 840–846.

[163] X. Wang, K. Tieu, and E. Grimson, *Learning Semantic Scene Models by Trajectory Analysis*. Berlin, Germany: Springer-Verlag, 2006, pp. 110–123.

[164] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Abnormal event detection from surveillance video by dynamic hierarchical clustering," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 5, pp. V-145–V-148.

[165] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anomalous event detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1544–1554, Nov. 2008.

[166] H. Huang, Z. Cai, S. Shi, X. Ma, and Y. Zhu, "Automatic detection of vehicle activities based on particle filter tracking," in *Proc. ISCSCT*, 2009, vol. 2, no. 2, p. 2.

[167] M. Pucher *et al.*, "Multimodal highway monitoring for robust incident detection," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2010, pp. 837–842.

[168] C.-T. Hsieh, S.-B. Hsu, C.-C. Han, and K.-C. Fan, "Abnormal event detection using trajectory features," *J. Inf. Technol. Appl.*, vol. 5, no. 1, pp. 22–27, 2011.

[169] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008.

[170] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *Proc. IEEE Appl. Imagery Pattern Recog. Workshop*, 2008, pp. 1–8.

[171] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[172] W. Hu, D. Xie, T. Tan, and S. Maybank, "Learning activity patterns using fuzzy self-organizing neural network," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 3, pp. 1618–1626, Jun. 2004.

[173] D. Vasquez and T. Fraichard, "Motion prediction for moving objects: A statistical approach," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, vol. 4, pp. 3931–3936.

[174] S. Atev, G. Miller, and N. P. Papanikolopoulos, "Clustering of vehicle trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 647–657, Sep. 2010.

[175] S. Gong and T. Xiang, "Recognition of group activities using dynamic probabilistic networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 742–749.

[176] Y. Xiong and D.-Y. Yeung, "Mixtures of ARMA models for model-based time series clustering," in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 717–720.

[177] M. Vlachos, G. Kollios, and D. Gunopulos, "Discovering similar multidimensional trajectories," in *Proc. Int. Conf. Data Eng.*, 2002, pp. 673–684.

[178] F. Bashir, A. Khokhar, and D. Schonfeld, "Real-time motion trajectory-based indexing and retrieval of video sequences," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 58–65, Jan. 2007.

[179] Y. Cao, C. Wang, L. Zhang, and L. Zhang, "Edgel index for large-scale sketch-based image search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 761–768.

[180] L. Olsen, F. F. Samavati, M. C. Sousa, and J. A. Jorge, "Sketch-based modeling: A survey," *Comput. Graph.*, vol. 33, no. 1, pp. 85–103, Feb. 2009.

[181] V. Saligrama, J. Konrad, and P.-M. Jodoin, "Video anomaly identification," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 18–33, Sep. 2010.

[182] Y.-K. Jung and Y.-S. Ho, "Traffic parameter extraction using video-based vehicle tracking," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 1999, pp. 764–769.

[183] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proc. IEEE*, vol. 89, no. 10, pp. 1456–1477, Oct. 2001.

[184] H. Aghajan and A. Cavallaro, *Multi-Camera Networks: Concepts and Applications*. Amsterdam, The Netherlands: Elsevier, 2009.

[185] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: A review," *Proc. Inst. Elect. Eng.—Vis. Image Signal Process.*, vol. 152, no. 2, pp. 192–204, Apr. 2005.

[186] X. Wang, K. Tieu, and E. Grimson, "Correspondence-free activity analysis and scene modeling in multiple camera views," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 56–71, Jan. 2010.

[187] M. Vazirgiannis and O. Wolfson, "A spatiotemporal model and language for moving objects on road networks," in *Advances in Spatial and Temporal Databases*. Berlin, Germany: Springer-Verlag, 2001, pp. 20–35.

[188] L. Speicvcys, C. S. Jensen, and A. Kligys, "Computational data modeling for network-constrained moving objects," in *Proc. Int. Symp. Adv. Geograph. Inf. Syst.*, 2003, pp. 118–125.

[189] J. Gudmundsson, M. van Kreveld, and B. Speckmann, "Efficient detection of motion patterns in spatio-temporal data sets," in *Proc. Annu. ACM Int. Workshop Geograph. Inf. Syst.*, 2004, pp. 250–257.

[190] P. Kalnis, N. Mamoulis, and S. Bakiras, "On discovering moving clusters in spatio-temporal data," in *Advances in Spatial and Temporal Databases*. Berlin, Germany: Springer-Verlag, 2005, pp. 364–381.

[191] J. Gudmundsson and M. van Kreveld, "Computing longest duration flocks in trajectory data," in *Proc. Annu. ACM Int. Symp. Geograph. Inf. Syst.*, 2006, pp. 35–42.

[192] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen, "Discovery of convoys in trajectory databases," *Proc. VLDB Endowment*, vol. 1, no. 1, pp. 1068–1080, Aug. 2008.

[193] L.-A. Tang *et al.*, "On discovery of traveling companions from streaming trajectories," in *Proc. IEEE Int. Conf. Data Eng.*, 2012, pp. 186–197.

[194] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, "On discovery of gathering patterns from trajectories," in *Proc. IEEE ICDE*, 2013, pp. 242–253.

[195] Z. Li, B. Ding, J. Han, and R. Kays, "Swarm: Mining relaxed temporal moving object clusters," *Proc. VLDB Endowment*, vol. 3, no. 1/2, pp. 723–734, Sep. 2010.

[196] C. S. Jensen, D. Lin, and B. C. Ooi, "Continuous clustering of moving objects," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 9, pp. 1161–1174, Sep. 2007.

[197] Z. Li, J.-G. Lee, X. Li, and J. Han, "Incremental clustering for trajectories," in *Database Systems for Advanced Applications*. Berlin, Germany: Springer-Verlag, 2010, pp. 32–46.

[198] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: A partition-and-group framework," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2007, pp. 593–604.

[199] P. Yan, "Spatial-temporal data analytics and consumer shopping behavior modeling," Ph.D. dissertation, Univ. Arizona, Tucson, AZ, USA, 2010.

[200] G. Gidofalvi and T. B. Pedersen, "Mining long, sharable patterns in trajectories of moving objects," *Geoinformatica*, vol. 13, no. 1, pp. 27–55, Mar. 2009.

[201] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proc. ACM SIGKDD Int. Conf. Mining*, 2007, pp. 330–339.

[202] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proc. ACM SIGKDD Int. Conf. Mining*, 2012, pp. 195–203.

[203] P. Laube, S. Imfeld, and R. Weibel, "Discovering relative motion patterns in groups of moving point objects," *Int. J. Geograph. Inf. Sci.*, vol. 19, no. 6, pp. 639–668, Jul. 2005.

[204] Y. Zheng and X. Zhou, *Computing With Spatial Trajectories*. New York, NY, USA: Springer-Verlag, 2011.

[205] A. Prasad Sistla, O. Wolfson, S. Chamberlain, and S. Dao, "Modeling and querying moving objects," in *Proc. Int. Conf. Data Eng.*, 1997, pp. 422–432.

[206] C. C. Aggarwal and D. Agrawal, “On nearest neighbor indexing of nonlinear trajectories,” in *Proc. ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2003, pp. 252–259.
[207] Y. Tao, C. Faloutsos, D. Papadias, and B. Liu, “Prediction and indexing of moving objects with unknown motion patterns,” in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2004, pp. 611–622.
[208] Y. Cai and R. Ng, “Indexing spatio-temporal trajectories with Chebyshev polynomials,” in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2004, pp. 599–610.
[209] Y. Tao, D. Papadias, J. Zhai, and Q. Li, “Venn sampling: A novel prediction technique for moving objects,” in *Proc. Int. Conf. Data Eng.*, 2005, pp. 680–691.
[210] J. Krumm and E. Horvitz, “Predestination: Inferring destinations from partial trajectories,” in *Ubiquitous Computing*. Berlin, Germany: Springer-Verlag, 2006, pp. 243–260.
[211] S.-W. Kim *et al.*, “Path prediction of moving objects on road networks through analyzing past trajectories,” in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2007, pp. 379–389.
[212] D. Tiesyte and C. S. Jensen, “Similarity-based prediction of travel times for vehicles traveling on known routes,” in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2008, p. 14.
[213] R. Simmons, B. Browning, Y. Zhang, and V. Sadekar, “Learning to predict driver route and destination intent,” in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, 2006, pp. 127–132.
[214] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, “Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior,” in *Proc. Int. Conf. Ubiquitous Comput.*, 2008, pp. 322–331.
[215] L. Chen, M. Lv, and G. Chen, “A system for destination and future route prediction based on trajectory mining,” *Pervasive Mobile Comput.*, vol. 6, no. 6, pp. 657–676, Dec. 2010.
[216] J.-G. Lee, J. Han, X. Li, and H. Cheng, “Mining discriminative patterns for classifying trajectories on road networks,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 5, pp. 713–726, May 2011.
[217] W. Mathew, R. Raposo, and B. Martins, “Predicting future locations with hidden Markov models,” in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 911–918.
[218] A. Y. Xue *et al.*, “Destination prediction by sub-trajectory synthesis and privacy protection against such prediction,” in *Proc. IEEE Int. Conf. Data Eng.*, 2013, pp. 254–265.
[219] E. M. Knorr, R. T. Ng, and V. Tucakov, “Distance-based outliers: Algorithms and applications,” *VLDB J.*, vol. 8, no. 3/4, pp. 237–253, Feb. 2000.
[220] J.-G. Lee, J. Han, and X. Li, “Trajectory outlier detection: A partition-and-detect framework,” in *Proc. IEEE Int. Conf. Data Eng.*, 2008, pp. 140–149.
[221] X. Li, J. Han, S. Kim, and H. Gonzalez, “ROAM: Rule-and motif-based anomaly detection in massive moving object data sets,” in *Proc. SIAM Int. Conf. Data Mining*, 2007, vol. 7, pp. 273–284.
[222] X. Li, Z. Li, J. Han, and J.-G. Lee, “Temporal outlier detection in vehicle traffic data,” in *Proc. IEEE Int. Conf. Data Eng.*, 2009, pp. 1319–1322.
[223] Y. Bu, L. Chen, A. W.-C. Fu, and D. Liu, “Efficient anomaly monitoring over moving object trajectory streams,” in *Proc. ACM SIGKDD Int. Conf. Mining*, 2009, pp. 159–168.
[224] Y. Ge *et al.*, “Top-eye: Top-k evolving trajectory outlier detection,” in *Proc. ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 1733–1736.
[225] D. Zhang *et al.*, “iBAT: Detecting anomalous taxi trajectories from gps traces,” in *Proc. Int. Conf. Ubiquitous Comput.*, 2011, pp. 99–108.
[226] L. Sun *et al.*, “Real time anomalous trajectory detection and analysis,” *Mobile Netw. App.*, vol. 18, no. 3, pp. 341–356, Jun. 2013.
[227] C. Chen *et al.*, “Real-time detection of anomalous taxi trajectories from GPS traces,” in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Berlin, Germany: Springer-Verlag, 2012, pp. 63–74.
[228] L. X. Pang, S. Chawla, W. Liu, and Y. Zheng, “On detection of emerging anomalous traffic patterns using gps data,” *Data Knowl. Eng.*, vol. 87, pp. 357–373, Sep. 2013.
[229] B. Krogh, O. Andersen, and K. Torp, “Trajectories for novel and detailed traffic information,” in *Proc. ACM SIGSPATIAL Int. Workshop GeoStreaming*, 2012, pp. 32–39.
[230] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. V. D. Wetering, “Visual traffic jam analysis based on trajectory data,” *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013.
[231] L. O. Alvares, V. Bogorny, J. A. F. de Macedo, B. Moelans, and S. Spaccapietra, “Dynamic modeling of trajectory patterns using data mining and reverse engineering,” in *Proc. Int. Conf. Conceptual Modeling*, 2007, vol. 83, pp. 149–154.
[232] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, “A clustering-based approach for discovering interesting places in trajectories,” in *Proc. ACM Symp. Appl. Comput.*, 2008, pp. 863–868.
[233] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, “Mining interesting locations and travel sequences from gps trajectories,” in *Proc. Int. Conf. World Wide Web*, 2009, pp. 791–800.
[234] X. Li, J. Han, J.-G. Lee, and H. Gonzalez, “Traffic density-based discovery of hot routes in road networks,” in *Advances in Spatial and Temporal Databases*. Berlin, Germany: Springer-Verlag, 2007, pp. 441–459.
[235] Q. Li, Z. Zeng, B. Yang, and T. Zhang, “Hierarchical route planning based on taxi gps-trajectories,” in *Proc. Int. Conf. Geoinformatics*, 2009, pp. 1–5.
[236] S. Oh *et al.*, “A large-scale benchmark dataset for event recognition in surveillance video,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 3153–3160.
[237] i-LIDS, Imagery Library for Intelligent Detection Systems. [Online]. Available: http://www.ilids.co.uk
[238] Autoscope. [Online]. Available: http://www.autoscope.com
[239] Citilog, Bala Cynwyd, PA, USA. [Online]. Available: http://www.citilog.com
[240] Kapsch TrafficCom, Vienna, Austria. [Online]. Available: http://www.kapsch.net/en/ktc
[241] Traficon. [Online]. Available: http://www.traficon.com
[242] HIKVISION. [Online]. Available: http://www.hikvision.com/en/us
[243] DAHUA, Hangzhou, China. [Online]. Available: http://www.dahuasecurity.com
[244] eris. [Online]. Available: http://www.iteris.com
[245] SONY. [Online]. Available: http://www.sony.net
[246] On Semiconductor, Phoenix, AZ, USA. [Online]. Available: http://www.onsemi.com
[247] N. K. Kanhere, S. J. Pundlik, and S. T. Birchfield, “Vehicle segmentation and tracking from a low-angle off-axis camera,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 1152–1157.
[248] X. Wang, T. X. Han, and S. Yan, “An HOG-LBP human detector with partial occlusion handling,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 32–39.
[249] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah, “Part-based multiple-person tracking with partial occlusion handling,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 1815–1821.
[250] S. Tang, M. Andriluka, and B. Schiele, “Detection and tracking of occluded people,” in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.
[251] K. Minoura and T. Watanabe, “Driving support by estimating vehicle behavior,” in *Proc. Int. Conf. Pattern Recog.*, Nov. 2012, pp. 1144–1147.
[252] G. Kogut and M. M. Trivedi, “A wide area tracking system for vision sensor networks,” in *Proc. 9th World Congr. Intell. Transp. Syst.*, Oct. 2002, pp. 1–11.
[253] B. T. Morris, C. Tran, G. Scora, M. M. Trivedi, and M. J. Barth, “Real-time video-based traffc measurement and visualization system for energy/emissions,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1667–1678, Dec. 2012.

Bin Tian received the B.S. degree from Shandong University, Jinan, China, in 2009. He is currently working toward the Ph.D. degree in control theory and control engineering with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include intelligent transportation systems, pattern recognition, and computer vision.

Brendan Tran Morris received the B.S. degree from the University of California, Berkeley, CA, USA, in 2002 and the Ph.D. degree from the university of California, San Diego, CA, in 2010.

He is currently an Assistant Professor of electrical and computer engineering and the founding Director of the Real-Time Intelligent Systems Laboratory with the University of Nevada, Las Vegas, NV, USA. He and his team researched on computationally efficient systems that utilize computer vision and machine intelligence for situational awareness and scene understanding.

Dr. Morris serves as an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (ITS) and the IEEE ITS MAGAZINE and is on the Board of Governors for the IEEE ITS Society (IEEE ITSS). He will serve as a Program Chair for the 2014 ITS Conference and 2016 Intelligent Vehicles Workshop. His dissertation research on "Understanding Activity from Trajectory Patterns" was awarded the IEEE ITSS Best Dissertation Award in 2010.

Ming Tang (M'06) received the B.S. degree in computer science and engineering and the M.S. degree in artificial intelligence from Zhejiang University, Hangzhou, China, in 1984 and 1987, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Chinese Academy of Sciences, Beijing, China, in 2002.

He is currently with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His current research interests include visual object detection, tracking, machine learning, and intelligent transportation.

Yuqiang Liu received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2010. He is currently working toward the Ph.D. degree in control theory and control engineering at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include intelligent transportation systems, graphical models, and computer vision.

Yanjie Yao received the B.S. degree from China University of Geosciences, Beijing, China, in 2011. She is currently working toward the Ph.D. degree in control theory and control engineering with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Her research interests include intelligent transportation systems, image processing, and computer vision.

Chao Gou received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. He is currently working toward the Ph.D. degree in control theory and control engineering at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include intelligent transportation systems, pattern recognition, and image processing.

Dayong Shen received the M.Sc. degree in system engineering from the National University of Defense Technology, Changsha, China, in 2013. He is currently working toward the Ph.D. degree at the College of Information System and Management, National University of Defense Technology.

His research work has been published in journals such as*Journal of Information and Computational Science*. His research interests include web data mining, social network analysis and machine learning.

Shaohu Tang received the B.Eng. degree in electronic information science and technology from Shandong Agricultural University, Tai'an, China, in 2009, and the M.Eng. degree in detection technology and automatic equipment from North China University of Technology, Beijing, China, in 2013. He is currently working toward the Ph.D. degree at Beijing Key Laboratory of Urban Intelligent Traffic Control Technology, School of Mechanical and Electronic Engineering, North China University of Technology.

His research interests include traffic control, intelligence algorithm, and traffic modeling.