A Survey of Vision-Based Traffic Monitoring of Road Intersections

Sokèmi René Emmanuel Datondji, Yohan Dupuis, Peggy Subirats, and Pascal Vasseur

***Abstract*—Visual surveillance of dynamic objects, particularly vehicles on the road, has been, over the past decade, an active research topic in computer vision and intelligent transportation systems communities. In the context of traffic monitoring, important advances have been achieved in environment modeling, vehicle detection, tracking, and behavior analysis. This paper is a survey that addresses particularly the issues related to vehicle monitoring with cameras at road intersections. In fact, the latter has variable architectures and represents a critical area in traffic. Accidents at intersections are extremely dangerous, and most of them are caused by drivers' errors. Several projects have been carried out to enhance the safety of drivers in the special context of intersections. In this paper, we provide an overview of vehicle perception systems at road intersections and representative related data sets. The reader is then given an introductory overview of general vision-based vehicle monitoring approaches. Subsequently and above all, we present a review of studies related to vehicle detection and tracking in intersection-like scenarios. Regarding intersection monitoring, we distinguish and compare roadside (pole-mounted, stationary) and in-vehicle (mobile platforms) systems. Then, we focus on camera-based roadside monitoring systems, with special attention to omnidirectional setups. Finally, we present possible research directions that are likely to improve the performance of vehicle detection and tracking at intersections.**

***Index Terms*—Omnidirectional vision, vehicle detection, vehicle tracking, behavior analysis, intersection monitoring.**

I. Introduction

VISUAL surveillance is widely used in many areas often for safety reasons. Several projects have led to important advances for vision-based traffic monitoring applications. In 1986, the European Research Program PROMETHEUS [1] was launched by the European automotive industry. It involved more than thirteen vehicle manufacturers and several research institutes from nineteen countries. The objectives of this pioneer project were to reduce road fatalities and improve traffic efficiency [2]. Later, the project VSAM was launched by the Defense Advanced Research Projects Agency, with the objective to develop an automated video understanding technology for use in future urban and battlefield surveillance applications [3]. Within this framework, it was reported an end-to-end testbed system demonstrating a wide range of advanced surveillance techniques such as real-time moving object detection and tracking from stationary and moving camera platforms or active camera control and multi-camera cooperative tracking. About two decades after these pioneer projects, the cooperative effort remained active with new European frameworks involving visual monitoring systems for intelligent transportations and road safety. As an example, the project ADVISOR [4] was successfully carried out in the early 2000s with the goal to develop a monitoring system for public transportation, in order to detect abnormal behaviors of users [5], [6]. However, despite the remarkable progress and efforts achieved by researchers, enhancing the safety of drivers is still a challenging issue, especially at road intersections.

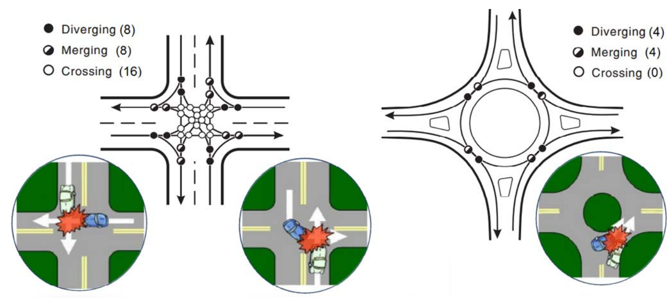


Fig. 1. Simplified illustrations of potential vehicle conflict points at a basic roundabout (8) versus a four-leg intersection (32) [7].

Intersection safety is a critical worldwide issue. In fact, accidents at intersections represent an important cause of road fatalities [8]. Intersections are particularly dangerous compared to highways because of their architectures which introduce several conflicts nodes (Fig. 1). Statistics from the U.S. Department of Transportation reveal that between 1998 and 2007, the number of fatalities at intersections exceeded 90,000 [9]. According to the European Road Safety Observatory, more than 62,000 people were killed in traffic accidents at intersections between 1997 and 2006 [10], [11]. Moreover, the proportion of fatalities in intersection accidents in EU throughout the decade 2000–2010, remained slightly equal to 20% of all cases [12]. More recently in EU, in 2013, more than 5.000 people were killed in road traffic accidents at intersections [13]. Cars, two-wheelers and pedestrians are particularly exposed to accidents at intersections [13]. Weather conditions are not a major cause

Manuscript received July 3, 2015; revised December 9, 2015; accepted February 3, 2016. Date of publication April 22, 2016; date of current version September 30, 2016. This work was supported by Région Haute Normandie and CEREMA. The Associate Editor for this paper was S. S. Nedevschi.

S. R. E. Datondji is with the Department of Infrastructure and Multimodal Transportations, Centre d'Études et d'Expertise sur les Risques, l'Environnement la Mobilitè et l'Aménagemant (CEREMA), 69674 Bron, France, and also with the Intelligent Transportation Systems, LITIS, Computer Science, Information Processing, and Systems Laboratory, 76821 Mont-Saint-Aignan, France (e-mail: rene-e.datondji@cerema.fr).

Y. Dupuis and P. Subirats are with the Department of Infrastructure and Multimodal Transportations, Centre d'Études et d'Expertise sur les Risques, l'Environnement la Mobilitè et l'Aménagemant (CEREMA), 69674 Bron, France (e-mail: yohan.dupuis@cerema.fr; peggy.subirats@cerema.fr).

P. Vasseur is with the Department of Intelligent Transportation Systems, LITIS, Computer Science, Information Processing, and Systems Laboratory, 76821 Mont-Saint-Aignan, France (e-mail: pascal.vasseur@univ-rouen.fr).

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.2016.2530146

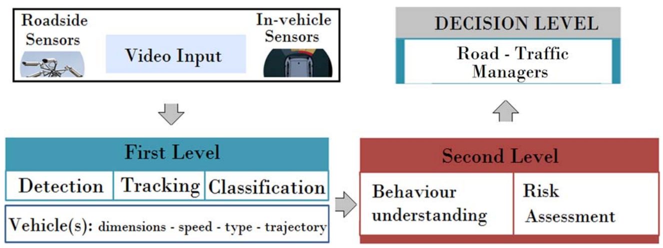


Fig. 2. Classic steps in video monitoring. First, vehicles are detected, tracked, and sometimes classified. Second, outputs of the previous step are used to understand vehicle behavior and evaluate the risk level.

of fatalities at intersections; because they affect accidents which occur away from intersections in a similar way [10]. However, lighting conditions represent a major factor as almost a quarter of fatalities at intersections happens during night time [10]. In general, regardless of the geometry of intersections or the meteorologic conditions, human decisions remain one of the most critical factors. In fact, more than 80% of accidents at intersections are caused by driver errors [8], [14]. In order to reduce by half the number of road deaths by 2020 [15], it is therefore necessary to develop innovative vehicle monitoring systems especially at intersections.

Vehicle monitoring consists in two levels of interpretation (Fig. 2). The first level consists in the actual scene modeling, vehicle detection and tracking. The output of this level provides data such as positions, speeds or classes of vehicles. In the literature, several papers have reported good performance for such tasks [16], [17]. The second level of interpretation consists in analyzing the interactions between vehicles to evaluate the risk-level [18]–[20]. This level allows to perform tasks such as predictions of specific behaviors, using the outputs of the previous level. Behavior interpretation has been actively investigated, not only to detect and analyze abnormal maneuvers, but also to prevent dangerous situations and anticipate conflicts [21], [22].

Vehicle monitoring at intersections is a special case which brings up several challenges. Vehicles at intersections can have variable and abrupt motions from different entry points. They can also occlude one another or be occluded by the infrastructure. Two categories of vehicle sensing technologies are generally used for monitoring vehicles at intersections: active sensors (RADAR, LIDAR) and passive sensors (Cameras - monocular vision, stereo vision, wide angle or omnidirectional vision). In this paper, we mainly discuss camera-based systems. In the recent literature, most intersection-related systems were developed for addressing collision avoidance and behavior analysis [23]. We can classify them into two categories: roadside and in-vehicle systems.

Roadside systems for intersections monitoring (Table I) are stationary platforms generally consisting in pole-mounted cameras or cameras placed on elevated buildings and connected to a central processing unit [24], [25]. Today, cameras are cheaper, smaller and smarter [26]. In addition, the rising power of processors, as well as the emergence of new generation of embedded architectures allowing real-time implementations, have spawned a great interest for camera-based systems [27]. At road intersections, most of these systems require one or multiple cameras to be mounted at highly elevated positions, which can be a major drawback for the deployment. Single-camera based systems are generally preferred to monitor intersections. Most of the works on multiple cameras-based systems even treat the information of each camera independently, and then perform a high-level fusion [28], [29]. In addition, in most cases important preprocessing steps, such as intrinsic and extrinsic camera calibration, are necessary before further traffic analysis. Besides, despite the increasing use of omnidirectional cameras in general ITS systems [30], they have been less exploited for roadside intersection monitoring systems. Only few representative studies have used wide field of view cameras (Table II) such as catadioptric [31] and dioptric [32], [33] vision sensors. However, these recent studies have shown the worthiness of omnidirectional cameras at intersections and have introduced several challenges. Despite the fact that they can enable to monitor an entire intersection, omnidirectional images involve the necessity to analyze important amount of visual data at various scales as the vehicles moves through the scene.

To date, the latest research works on roadside intersection monitoring [32], [33] have led to the development of commercial applications. The company GRIDSMART claim to built intersections monitoring systems, upon the foundations of simplicity, flexibility and transparency, with the goal to empower traffic managers. They pointed out the necessity for traffic professionals to get started easily with new technologies and use their own computers to design, organize, configure and manage intersections [34]. This is a perfect illustration of the growing challenges related to non-intrusive detection and tracking technologies for intersections monitoring with stationary omni-vision systems. Thus, the problem of monitoring traffic participants at intersections has moved toward another level. Nevertheless, the research is still opened, because to the best of our knowledge none of the existing systems are fully automated, can provide either complete reliability or robustness. In this survey, the case of roadside intersection monitoring with omnidirectional cameras will be given a special focus.

In-vehicle systems for intersections monitoring (Table III) are also generally based on cameras, often stereo setups with short baselines, installed or embedded in mobile platforms [36], [37]. Today, new generations of in-vehicle driving support systems such as cooperative advanced safety systems, based on sensor fusion, are being actively studied for applications at intersections (Fig. 3) [38]. In the European project INTERSAFE [39], a mobile platform for accident detection at road intersections was developed combining a wide range of active and passive sensors. Recently, within the framework of the project Ko-PER [40], [41], a system combining laser scanners with low and high resolution cameras has been used to gather classification information about vulnerable road users. Thus, recent research works about intersections safety systems aims not only to reduce the risks of accidents, but also to provide solutions for reliable data collection, as well as innovative technologies for drivers behavior analysis.

Interesting reviews have been proposed about traffic monitoring and its applications in general [17], [42]–[44]. But to the best of our knowledge, none of these surveys paid a particular attention to intersection-like scenarios. In the paper, the unique challenges of intersections are explicitly discussed and analyzed. We must say that this survey is directed toward camera-based systems, and a special discussion will be given about roadside omnidirectional vision monitoring applications.

TABLE I
SELECTED REPRESENTATIVE ROADSIDE SYSTEMS FOR TRAFFIC MONITORING AT INTERSECTIONS

Roadside works	Sensor setup	Method	Description
(2000) Kamijo et al. [66]	Monocular vision	Frame differencing by Spatio-Temporal Markov Random Field	The algorithm was evaluated on real traffic images, with no assumption of any physical models like shape or textures. They performed multiple vehicle detection and tracking at intersections with occlusion and clutter effects at the success rate of 93%–96%
(2004) Messelodi et al. [25]	Monocular vision	Frame differencing	Motion segmentation by estimating the moving map. The performance of detection and classification is over 90 % in good weather. The system does not work during night.
(2005) Arvind M. [48] [45]	Monocular vision (visible and infrared)	Adaptive background subtraction	Region-based segmentation by iterative thresholding, for real-time vehicle trajectory estimation at rural intersections. At night, infrared cameras are used to detect vehicle wheels.
(2005) Heikki et al. [162]	Monocular vision	Adaptive background subtraction	Velocity profiles of each vehicle is computed from a series of consecutive images. The system cannot deal with cast shadows or occlusions.
(2006) Saunier et al. [77]	Monocular vision	Motion-based segmentation (feature vertices)	Points tracks (temporal series of coordinates) are grouped together to generate vehicle hypotheses. Authors reported 88.4% of occluded vehicles detected. They explained that most error are introduced by feature over-grouping
(2014) Furuya et al. [67]	Monocular vision	Background substraction + Motion segmentation	Motion segmentation by KLT and corner feature grouping based on similar velocity profiles. Vehicle detection with 56% accuracy. The large error is attributed to the large perspective deformation (Figure 20)

TABLE II
OMNI-VISION BASED SYSTEMS FOR INTERSECTION MONITORING

Omni-vision systems	Description
(2006) Lee et al. [32]	Tracking in image domain by particle filter and motion dynamics – Important failure caused by appearance changes. Not robust to various weather conditions.
(2009) Ghorayeb et al. [31]	Pole-mounted optimized catadioptric camera providing 90% of useful image area – Detection by background subtraction
(2011) Gee et al. [33]	Pole-mounted at variable elevated positions – Detection by background subtraction – 3D-model based tracking
(2015) Wang et al. [74]	Fisheye camera at variable elevated positions – Commercial application – Detection by background subtraction and motion-based verification – Tracking by KLT algorithm with concept of grafting and identity-appearance under constrained motion.

TABLE III
SELECTED REPRESENTATIVE IN-VEHICLE SYSTEMS FOR TRAFFIC MONITORING AT INTERSECTIONS

In-vehicle Works	Sensor setup	Method	Description
(2009) Barth et al. [37]	Stereo vision	Optical flow - Motion cues for clustering 6D points	Vehicles are represented as rigid 3D points clouds, which are grouped based on common motion. Real-time dense stereo disparity maps provide compact stixel world representation.
(2011) Aycard et al. [39]	Stereo vision	Sparse optical flow of corners features	A two-level architecture providing 6D-point (3D motion and 3D position) information for obstacle detection
(2012) Muffert et al. [165]	Stereo vision	Dynamic Stixels.	They are extracted from a stereo image using SGM along with motion data (optical flow or 6D-vision estimations)

Our survey paper is organized as follows. In Section II, we start by giving a brief overview of vehicle sensing technologies used at intersections and present different datasets. In Section III, we present a short introductory review of general vehicle detection and tracking approaches. Then in Section IV, we tackle the context of intersection monitoring. In this section, we detail the particular challenges of vision-based vehicle detection and tracking at intersections. Then, we discuss roadside and in-vehicle solutions, around three subsections: monocular vision, stereo vision, omnidirectional vision. In Section V, we present a summary, a comparative analysis between in-vehicle and roadside systems for intersection monitoring, as wells as a discussion of future trends. Finally section VI concludes and presents the challenges and perspectives, especially in the context of roadside vision-based traffic diagnosis of intersections with omnidirectional cameras.

Fig. 3. Intersection safety systems and trends. New generations of driving support systems for infrastructure to vehicle communications and cooperative advanced safety systems are being studied [35].

Fig. 4. MIT data set [51], [60].

II. Sensing Vehicles at Intersections

In this section, we present the different sensing technologies used for monitoring vehicles at road intersections. We can distinguish between active and passive sensors. Active sensors such as RADAR and LIDAR, consist of a source that emits waves onto the target. The waves are reflected back by the object towards a receptor. Unlike active sensors, passive sensors only work as receptors that measure information either emitted or reflected by the object. In other words, sensors such as cameras, that use external energy sources to observe an object are passive. After a discussion of these sensing solutions, we will present several datasets that can be used for intersection monitoring studies.

A. Sensing Technologies for Intersection Monitoring

1) RADAR: stands for RAdio Detection And Ranging, and has been used to detect objects in intersections-like scenarios [23]. In [39], four short range RADARs (SRR) placed at each corner of the vehicle and a long range RADAR (LRR) in front of the vehicle were used (Fig. 22(a)). RADAR can help determine the range, the altitude, direction, or speed of an object. Motion and dynamics parameters are the fundamental cues for RADAR-based vehicle detection and tracking systems. However it might cause false detections between different types of objects having the same motion model. As a consequence, RADAR data can be very noisy and require extensive post-refinements [39]. Though RADAR designed for traffic monitoring have a very narrow field of view, they can be robust to difficult weather and illumination [45].

2) LIDAR: it stands for Light Detection and Ranging. The use of this sensing technology for traffic applications and vehicle monitoring has increased thanks to the reducing cost of the technology. It is an optical remote sensing technology which measures properties of scattered laser-wave to find range and additional parameters of a distant target. The measurement is based on the estimation of time-of-flight of a laser wave. Thus, it can be used to detect stationary as well as moving objects. In [39], laser beams are classified in a local grid map as static or dynamic in order to segment moving objects from static ones. In [46], single-row laser scanners data were used to perform trajectory and behavior analysis of vehicles passing at an intersection. The system classifies large amounts of trajectories based on a group of route models built from trajectory clustering. In [47], 3D clouds acquired from a dense beam scanning LIDAR mounted on the roof of an autonomous vehicle are used to detect when the vehicle reach an intersection. LIDAR has proved its efficiency, providing cleaner data than RADAR, but is sensitive to the environment and the weather.

3) Cameras: are widely used for traffic monitoring at intersections and provide rich visual information. Objects are visible with the camera because of the light reflexion from their surface onto the vision-sensor. A point in 3D world reference is mapped onto the image plane reference, into a 2D point via a projection matrix. For daylight operations, visible light cameras can be used. However by night or during difficult weather, a visible light camera is unlikely to meet good performance needs over extended periods. The use of infrared cameras can be a good solution for night-time vision, as for instance in the night the temperatures of the wheels of the vehicle are higher than the isotherm temperature [48].

Camera networks can offer several advantages such as: acquiring richer data, solving occlusion issues, enabling redundancy. However, deploying a network required to take into account critical points such as: mobility, power consumption, and compulsory spatial and temporal calibration [49]. In the context of intersections, camera networks are hardly used.

Omnidirectional cameras have been widely used in many areas as they possess wider field of view (FOV) than conventional cameras. However, they have been hardly used for intersection monitoring, as they introduce several challenges in this context such as: the real-time analysis of large image provided in a single shot, with highly distorted objects at variables scales [33]. Moreover, existing roadside systems with omnidirectional cameras cannot be easily adapted to different types of intersections [31] without important civil engineering.

B. Datasets for Traffic Analysis at Intersections

The interest for traffic monitoring at intersections has kept increasing over the last decade, with an emphasis on environment modeling and vehicular behavior analysis. The cooperative effort contributes to the exchange of data and the advances in research. This has led to the emergence of challenging datasets for evaluation and benchmarking.

1) MIT (Fig. 4): the traffic dataset from the Massachusetts Institute of Technology, is for research on activity analysis of crowded scenes [50]. It holds nineteen minutes of raw video recorded by a stationary camera, including vehicle traffic at a four-way signalized intersections [51].

2) NGSIM (Fig. 5): the next-generation simulation program was initiated to develop a core of open behavioral models in support of microscopic modeling and traffic simulation [52], [53]. It includes validation data that have been used more recently for learning traffic behaviors and patterns [54]–[57]. The

Fig. 5. NGSIM data set [52].

Fig. 6. CBSR data set [58].

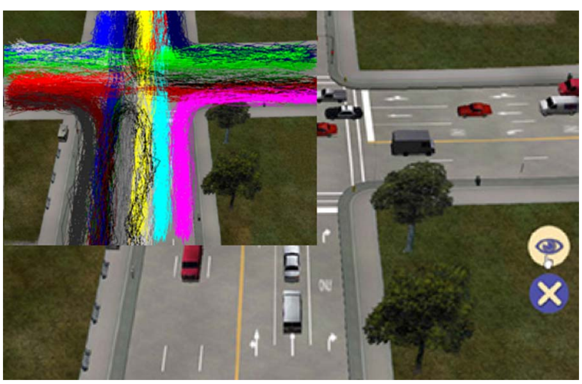


Fig. 7. CVRR data set [20].

overhead intersection cameras provide multi-view raw video data of intersections, with detailed vehicle trajectories, as well as supporting behavioral data.

3) CBSR (Fig. 6): the traffic dataset from the Center for Biometrics and Security Research, provides a single view of a complex intersection, with several possible entry and exit combinations. It has been essentially used for motion patterns analysis to predict abnormal behavior [58].

4) CVRR (Fig. 7): the traffic dataset from the Computer Vision and Robotics Research Lab, provides data for benchmarking unsupervised trajectory-based activity analysis algorithms: clustering, classification, prediction, and abnormality detection. This dataset consists of a simulated intersection, a real highway, as well as two indoor omnidirectional camera videos. However, the provided trajectories contain only spatial information. Besides, true activity labels are provided as well as full tracks based abnormality label and frame-by-frame unusual event information [20].

5) QMUL (Fig. 8): the dataset from Queen Mary University of London is specifically intended for activity analysis and behavior understanding [59], [60]. It contains one hour of traffic video collected at a busy intersection with ground truth.

Fig. 8. QMUL data set [60].

Fig. 9. Ko-PER data set [41].

Fig. 10. KIT data set [61].

6) Ko-PER (Fig. 9): the Ko-PER project datasets are meant to generate a comprehensive dynamic model of the ongoing traffic. The datasets consist of data from the laser-scanners network and cameras installed at a public intersection, as well as reference data and object labels. The datasets are shared for further research in the field of multi-object detection and tracking [41].

7) KIT (Fig. 10): the intersection traffic datasets from Karlsruhe Institute of Technology [61] provide eight videos recorded by a stationary calibrated camera at different location and in various conditions (rain, snow, fog). The dataset has been mainly used to study vehicle detection and tracking at intersections [62].

Fig. 11. Urban tracker [63].

8) Urban Tracker (Fig. 5): The Urban Tracker dataset, shared by researchers from the École Polytechnique de Montréal, focuses on traffic research applications. It provides recordings of the traffic scenes, meta-data, camera calibration, ground truth, protocols for comparing algorithms, software tools and libraries for reading the data [63] (see Fig. 11). There are four annotated videos sequences of intersections, recorded in different weather and lighting conditions (Sherbrooke, Rouen, St-Marc, René-Lévesque) from stationary cameras at different heights. This dataset is suitable for vehicle and pedestrian tracking at intersections, with several computer vision challenges.

III. Generalities of Vision-Based Vehicle Monitoring

Under this section, we review vision-based monitoring in the general sense. In the first subsection, we present the major challenges in this field. In the second and third subsections, we give an introductory review of vision-based vehicle detection and tracking, respectively.

A. Challenges of Vision-Based Vehicle Monitoring

There are several challenges regarding vision-based vehicle monitoring. First of all, the quality of the sensor itself (noise, vibration, image format and resolution, lighting and optics, color representation, computation power available [64], [65]) might affect the monitoring process. Besides, there are important initialization and steps (camera calibration, image region of interest definition, initial maps) which are context-based. After a discussion about general initialization and preprocessing issues, we present challenges intrinsically related to vision-based vehicle detection and tracking.

1) Initialization and Preprocessing Challenges: in many camera-based traffic monitoring systems, there are often compulsory and context-based initialization tasks, such as image formatting and rectification or initial map generation [66], [67]. For instance, in the particular case of omnidirectional cameras, image distortions and resolution are important issues for real-time systems. In fact, it is necessary to count a large number of pixels, to extract useful regions in the image, and monitor targets spanning nearly four order of magnitude [31]. Thus, in order to design a robust vision-based monitoring system for traffic applications, it is required to take into account several parameters and external factors.

Traffic monitoring systems require calibrated cameras to perform image to world mapping. Accurate calibration is necessary to estimate vehicle dimensions and motion parameters. It includes the estimation of camera intrinsic parameters on the one hand, and the extrinsic parameters on the other hand. While intrinsic calibration can be determined beforehand, extrinsic calibration need to be assessed once the cameras are installed [68], [69].

The calibration of roadside cameras mostly relies on geometric constraints and prior knowledge of the scene. This task is mainly based on the estimation of vanishing points (VPs), calculated with parallel lines extracted from specific landmarks such as lane marking. The number of vanishing points required for calibration depends on the measurement available on the environment [25], [68], [70]. As a consequence, their accuracy highly depends on the user inputs and assumptions. A detailed taxonomy of roadside camera calibration methods is proposed in [71]. These methods cannot be efficiently applied when well-visible patterns with parallel lines on the road are hardly available. As an alternative, motion-based VPs estimation has been proposed as a good solution, provided the assumption of straight and planar motion [72].

2) Vehicle Detection and Tracking Challenges: Vehicles have different characteristics such as size, color or shape. A fundamental question at initialization is the choice of an adequate and robust vehicle feature representation with respect to the application (blobs [73], set of points [74], or geometric models [75]). In general the vehicle representation defines the tracking strategy [76]. The most important goal is the ability to keep vehicle tracks active as long as possible in the scene. Reliable vehicle tracking is crucial, as it will allow posterior trajectory analysis and behavior recognition [5].

Through the literature, it appears that vehicle detection and tracking are somewhat related tasks. Some methods require vehicle to be detected first and then tracked, while other strategies use vehicle tracks as cues for detection. There are several issues that make vehicle detection and tracking challenging. The major challenges reported are: vehicle occlusion (vehicle-vehicle, infrastructure-vehicle), changes of vehicle perception (appearance, camera placement, distortion, size, scale [67]), sudden vehicle motion [77] or sudden change in the environment (illumination, weather, night time vision) [43], shadow detection and removal [78]–[80].

B. General Review of Vision-Based Vehicle Detection

Under this section, we present the general approaches for vehicle detection. Though many papers in the recent literature have reported good performance of vehicle detection, it remains a major issue due to the several challenges presented above. In general, vehicle detection is fully performed in two steps: finding foreground entities considered as vehicle hypothesis, and then verifying these candidates.

1) Vehicle Candidate Localization: it can require one or several frames. The different methods can be classified into four categories: background subtraction, model-based segmentation, feature-based segmentation, motion-based segmentation. The methods also vary with the system setups. In monocular vision, appearance-based cues are mostly used to obtain the vehicle hypothesis frame by frame. Adaptive motion models have also been often applied to differentiate vehicles from the background

[81], [82]. Whereas in stereo-vision, motion-based methods are the current trend. In this case, multi-view geometry enables direct measurements of 3D information. Stereo-vision has been efficiently used to separate obstacles from the free space, thanks to the disparity maps [83], dynamic occupancy grids [84], or inverse perspective mapping [85].

a) Background subtraction: it consists in extracting foreground objects from a single image by removing a reference background model [86], [87]. The major difference between background subtraction approaches in the literature, relies in the way to obtain the background. It can be either static, dynamic, statistical or adaptive background estimation. There are several problems related to this task, because of the difficulties to define boundaries between background and foreground [79]. The most popular background modeling method is the Gaussian Mixture Model (GMM) [88]. It consists in modeling pixel values over time by a weighted mixtures of Gaussian [89]. Later, Barnich Olivier *et al.* introduced a novel approach called: the universal video background subtraction (ViBe). The model consists in a set of observed pixel values [78]. The pixels are classified as background by thresholding the distance from a given pixel and all samples. ViBe algorithm incorporates a memoryless update policy and is resilient to noisy data.

b) Model-based segmentation: it consists in identifying possible foreground vehicles in an image by fitting vehicle 2D-predefined or 3D-projected shapes to image regions, without any knowledge of a background model [25], [90]. However, these direct matching approaches are unrealistic because it is impossible to have a model for all possible vehicles that may be present in the scene. The use of probabilistic frameworks [25], [91], [92] or motion information [62], [75], [93] along with vehicle models has been often used to reduce the matching search space.

c) Feature-based segmentation: searching by analyzing the geometry or appearance features is a common method for vehicle candidate localization in the foreground. Researchers have used texture [94], [95], color [96], shadow [97], [98] and geometric elements such as corners [99], and symmetry analysis [100], [101] or fusion of several cues [102]. A good vehicle feature should be able to capture the distinctive characteristics and be robust enough to small variations over different background conditions [103]. This is necessary in order not only to reduce the dimensionality of the data to be processed, but also the computation time. More recently, new trends of descriptive features have been used because they enable a more direct vehicle hypothesis and localization: SIFT [104], ASIFT [105], SURF [106], Histogram of Oriented Gradient (HOG) [107], [108], Gabor features [109], [110], Haar-like Wavelets [111].

d) Motion-based segmentation: consists in identifying moving items in the foreground by searching image regions having significant changes between successive frames [16]. It uses the pixel-level difference and the temporal information to extract moving regions. Motion-based hypothesis generation makes use of temporal information to detect vehicles, and obstacles by matching and grouping image pixels having the same motion characteristics over consecutive frames [99]. It can rapidly adapt to dynamic environments when temporal changes are important, but it may fail to extract all the representative pixels in complex scenes (appearance or scale change, shape change, variable motions, scale change, occlusion or clutter) [112]. Moreover, generating a displacement vector for each pixel is time consuming and computationally expensive. The discretization of the image gives better results and enables real time processing [106].

The optical flow is one of the most popular approach freely available [113]. It can be seen as the projection of a three-dimensional motion field onto the image plane. Dense stereo algorithms provide significantly more information on the 3D environment compared to sparse stereo methods. The gain in information and precision allows for improved scene reconstruction and object modeling. However, in order to reduce computation time, the sparse optical flow is preferred [114], as it considers only a sets of relevant pixels. The latter generally gives enough information to formulate the hypothesis about vehicles while being less sensitive to noise [115]–[117]. There has been several adaptations of the optical flow based on compact shapes such as stixels [118]–[120] in order to obtain a more comprehensive representation, but also to speed up the processing. Besides, optical flow can also be used for understanding complex road scenarios. Geiger *et al.* [121] introduced the object flow descriptor which works as an urban traffic classifier, in particular to detect when vehicles reach an intersection.

2) Vehicle Candidate Verification: Regardless of the approach selected for identifying potential vehicle candidates in the foreground, vehicle detection is completely achieved only after a verification procedure. The latter needs to be performed to discard false alarms, and ensure that the candidates are actually vehicles. We discuss vehicle verification techniques in two categories which are similarity estimation and discriminative classification. The former calculates correlation score between a given vehicle candidate and a given template, while the latter verifies the candidate after a learning process.

a) Similarity estimation: it consists in using predefined templates, and estimating their correlation between a vehicle candidate region. A similarity is expressed as the result of vehicle verification [122]. Once a vehicle hypothesis has been verified, it can be used dynamically as a new template if its correlation score or reliability exceeds a certain threshold [123]. Three-dimensional templates can be projected into 2D-templates and matched with images regions [90], along with probabilistic frameworks [92]. In [25], the convex hull for 3D vehicle models were generated in the image. The ratio between convex hull overlap of model and image normalized by the union of both areas generates a similarity score. Furthermore, 3D-models have been used as well to verify motion-based vehicle hypothesis [62], [93]. In [75], N. Buch presented a vehicle detection and classification system for urban traffic scenes in which: vehicles were detected in each frame using 3D models; then motion silhouettes were extracted and compared to a projected model silhouette, in order to identify the ground plane position and classes of vehicles. They tested the system successfully in three weather conditions and the full system including detection and classification for all data in various weather achieves a recall of 90.4% outperforming similar systems in the literature.

b) Discriminative classification: it consists in learning a decision boundary between two classes of features. The goal is to distinguish between non vehicles and vehicles objects. The classification is performed by first learning the appearance of a vehicle from a training dataset. The training is generally based on a supervised approach where a large set of labeled positive (vehicle) and negative (non-vehicle) images are used. Common techniques used in recent literature for vehicle verification are: Support Vector Machines (SVM) [124], Artificial Neural Networks (ANN) [125] and AdaBoost [126].

SVM (Support Vector Machines) classifiers are developed based on the statistical learning theory described in [124]. The idea is to map the training data of two object classes from the input space into a higher dimensional feature space. Then an optimal separating hyperplane with maximum margin is constructed in the feature space to separate the two classes [110]. Based on a set of trained orientated HOG features, [107] successfully classified vehicle by orientation, with 88% accuracy. In [127], Gabor features, which capture the local lines and edges information at different orientations and scales, were trained on the SVM classifier for vehicle detection. The evaluation results show that the classifier can achieve 94.5% detection rate. In [128], the combination of boosted Gabor features enabled to reach 96% detection rate. By combining Gabor and Legendre moment features for training an SVM classifier, [129] reported a better performance of 99% for vehicle detection. In order to reduce the dimensionality of the data, [130] applied Principal Components Analysis (PCA). The authors build into a new sub-space a specific vehicle feature vector. Doing so, they obtain an excellent performance on the SVM training with 95% accuracy.

Artificial Neural Network (ANN) classifiers are less used for vehicle detection in the late literature. In 2000, [131] proposed a vehicle classification system which calculates texture features using the co-occurrence matrix, within a classification scheme based on multilayer perceptron. More recently ANN were used as well for infrastructure detection [132] or for vehicle license plate recognition based on sliding concentric windows [133]. In [134], a probabilistic neural network framework has been proposed. The authors reported a maximum performance of 69.38% of successful automatic detection. They also claimed that the proposed approach has a substantially higher degree of performance, both qualitative and quantitative, than other state-of-the-art methods. However, ANN-based vehicle classification is less efficient than SVM-based methods.

AdaBoost [135] (Adaptive Boosting) classification was first used to perform face detection, with a set of Haar wavelet features; in a constructed cascade of increasingly more complex classifiers [136]. In [137], a boosted cascade of weak classifiers is used to analyze the redness of tail lights by night, in order to detect vehicles on the road. Using a richer set of Haar features, [138] were able to detect cars and buses with 71% detection rate. Adaboost-based classification schemes have been less studies in the recent literature for vehicle monitoring [139].

C. General Review of Vision-Based Vehicle Tracking

1) Vehicle Representation and Tracking Approaches: there are several approaches for vehicle tracking that have been discussed in the literature. Depending on the vehicle representation, which range from pixel level to object level, four general approaches can be distinguished: region-based tracking, contour-based tracking, feature-based tracking, model-based tracking.

a) Region-based tracking: regions or blobs can be defined as connected image parts with distinguishing common properties, such as intensity, color or texture statistics. Region or blob based tracking aims at tracking vehicles according to variations of the image regions [76]. An entire region associated to a given vehicle is tracked over time using appearance, geometrical properties as well as motion cues. Region-based tracking is computationally efficient and works well in free-flowing traffic. However, under congested traffic conditions, complex deformation or a cluttered background, vehicles can partially occlude one another instead of being spatially isolated [140].

b) Contour-based tracking: contour-based tracking algorithms represent objects by their contours, which are simply their boundary, and update these contours dynamically at every time increment. In fact, strong changes in image intensity generally occur at contours, which make it suitable for tracking purposes. Contour-based tracking algorithms provide more efficient description of vehicles than region-based algorithms: by reducing the computational time and complexity. However contour-based tracking does not perfectly solve occlusions. Even though vehicle contours can be extracted separately, a difficult task is to isolate vehicles being tracked if the contours of different vehicles are merged at some point. It requires an active track grouping and update policy [77].

c) Feature-based tracking: feature-based tracking use the principle that vehicles can be represented by a set of features, instead of an entire object. This refers to the group of methods that perform tracking by first extracting features in independent images and then matching the features over the frames. Features can be selected as representative parts of the vehicle, such as corners, lines, typical shapes. This technique is effective as long as the selected features are robust enough and can be distinguishable even if the vehicle is partially occluded at some point in the sequence. As suggested by [77], the feature-based tracking is suitable for daylight, twilight or night-time conditions, as well as different traffic conditions. But since a vehicle can have multiple features, the major problem resides in defining conditions that allow proper grouping [67] or clustering of those features from a given vehicle. The main cues for grouping are spatial proximity and common motion. Feature-based algorithms can adapt successfully and rapidly, allowing real-time processing and tracking of multiple vehicles in dense traffic. Moreover, for a feature-based tracking approach to be reliable, it must minimize the risk of mismatches through robust estimation algorithms.

d) Model-based tracking: this approach consists in matching a projected model onto the image, as the vehicle moves frame by frame. This allows to recover trajectories and models, as well as the pose of the vehicle with higher accuracy [141], [142]. In [141], a 3D cuboid is used as the vehicle model. The matching between the measurement and the model is performed by an intersection of rectangles in the birds-eye view, and a corner by corner matching in the image space. In [142], vehicles are modeled by a cloud of 3D points under the rigid body assumption. The vehicle model is updated by fusing stereo vision and tracking of image features. The major weakness is the need for accurate geometric object models. It is however unrealistic to have detailed models for all vehicles in the traffic, therefore additional cues are generally added upon initialization [62].

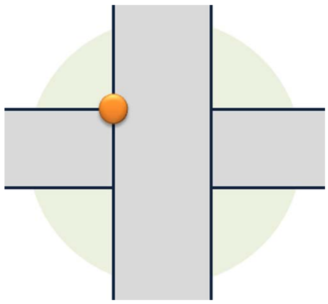


Fig. 12. Camera in a corner of the intersection.

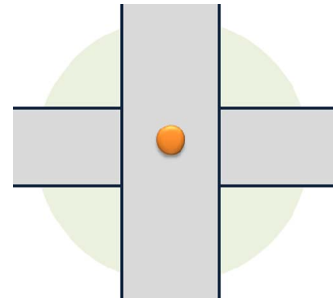


Fig. 13. Camera in the center of the intersection.

2) Popular Algorithms for Vehicle Tracking: they are powerful mathematical tools that can be either iterative or non-iterative. Iterative techniques can solve the correspondence problem between detected vehicles over frames; however without additional mechanisms they may not fulfill real-time purposes. Non-iterative approaches are used to overcome such limitations. Moreover, tracking can be applied based on single, or multiple hypothesis which are likely to improve the accuracy of the tracker at the cost of computational power [143]. Most of the trackers use the following steps: initialization and prediction; observation and data association; track update. We can distinguish between Matching and Bayesian tracking algorithms.

a) Matching algorithms: use image features to steer a tracking hypothesis iteratively and therefore tend to refine the state estimation until convergence. In general the goal is to align a given template within the image frames in order to calculate the displacement iteratively. The popular Kanade-Lucas-Tomasi tracker uses an appearance-based model on a template to track the object. The tracker is based on the early work of Lucas and Kanade [144] and was fully developed by Tomasi and Kanade [145]. KLT is a standard for vehicle tracking [77], and still often used in the recent literature [67] (Fig. 20), [72], [74] for complex road scenes.

b) Bayesian tracking algorithms: model a state and the related observations as two stochastic processes. The goal is to estimate the probability of the next state, given all previous measurements, based on a conditional probability density function. The Kalman filter, also known as linear quadratic estimator [146] is one of the most popular methods [25], [147]–[149]. Because real-life dynamic problems are often non-linear, there have been several variations of the traditional Kalman filter [150]: the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF) aim to address the issue of (highly) non-linear dynamic systems [151]. Another Bayesian framework is the particle filter introduced in [152]. It aims to generalize the problem of the Kalman filter to non-linear systems [153] and overcomes the constraint of a single Gaussian distribution of Kalman filters. This allows to model more complex distributions as well as nonlinear transformations of random variables. The filter can enable multiple hypotheses propagation between frames by modeling arbitrary probability density functions based on particle sampling. Several applications of the particle filter for vehicle tracking can be found in the recent literature [154]–[157]. Bayesian frameworks for tracking remain however the trend over the recent literature.

So far, we have presented vehicle detection and tracking in general traffic monitoring applications. Earlier in this survey, we have also introduced vehicle sensing technologies as well as representative datasets for vehicle monitoring studies at road intersections. In the next part, we will focus on the special case of intersections. We will present challenges brought up by intersection monitoring systems, then we will review in-vehicle and roadside vehicle detection and tracking by monocular, stereo, and omni-directional vision.

IV. Vision-Based Vehicle Monitoring at Road Intersections

A. Challenges Arising in the Context of Intersections

Meanwhile, vehicle monitoring for ITS has been an active research area for decades and achieved promising results, only a few studies have been attempted so far for intersections. Vehicle monitoring at intersections, either with roadside or in-vehicle systems, is particularly challenging for many reasons.

1) System Setup Challenges: these are essentially addressed to roadside systems, for which camera installation at the intersection needs to be flexible. Existing systems use generally a single camera looking at an intersection from variable positions (Figs. 12 and 13). Besides, vehicles need to be detected and tracked from many entry points [25], [32]. Moreover, it is also necessary that vehicles are visible from a relatively long distance, not just when they arrive exactly at the intersection; which is not always possible with traditional video-surveillance cameras. Thus, in order to observe the entire intersection in a single image, and monitor vehicles approaching on a long distance (Figs. 14 and 15), omnidirectional cameras have been recently used in this field [33], [74], [158]. However in all these works, cameras need to be pole-mounted at elevated positions and this task can be sometimes bulky [158]. Roadside camera systems should instead be designed to adapt to the infrastructure and must be easily installable. Thus, design challenges and constraints for camera-based roadside intersection monitoring are fundamentally: on the one hand, the easiness of deployment with flexibility to several types of intersections; on the other hand, the ability to monitor targets approaching from far distance, through the entire intersection.

2) Computer Vision Challenges: as many general vehicle monitoring applications, almost all the existing systems for intersections monitoring require user input at initialization. Besides, the extrinsic calibration of roadside cameras is still a major challenge [159]. For a single camera, it refers to recovering the orientation of the sensor with respect to the road or traffic stream. Whereas in case of multi-camera systems, extrinsic calibration refers instead to the process of recovering the transformations that relates cameras to one another in the network. As defined, the extrinsic calibration is not straightforward. The main reason for that is the large baseline, which does not allow the use of

Fig. 14. Example view from a fisheye camera in a corner of an intersection [32].

Fig. 15. Example view from a fisheye camera in the center of an intersection [33].

traditional calibration grids or patterns on the ground. To the best of our knowledge multiple-camera networks are hardly used for roadside intersection monitoring, though they could offer several advantages in order to accurately track vehicles which can be easily occluded or have sudden motions [160].

The complexity of vehicle monitoring at intersections is also amplified due to vehicle motions. In fact, as opposite to highways, vehicles at intersections can have variable, abrupt motions and can easily be occluded [77], [161]. Consequently vehicle tracking, trajectory prediction, as well as learning motion patterns in this context remain challenging and opened issues [21], [22], [58], [160].

B. Vision-Based Vehicle **Detection at Intersections**

1) ***Roadside*** Intersection Monitoring Systems:

a) ***Monocular vision****:* early studies suggested that automatic monitoring algorithms at intersections should be based on local analysis of individual vehicle behavior [66]. In this context, adaptive background subtraction methods have been often used to detect vehicles [73], [161], [162] (Fig. 17). This method has proven to be robust enough in case of illumination changes, camera noise, moving tree leaves, slow moving objects, but the performance is generally affected by shadows.

Messelodi *et al.* developed the system SCOCA which is a flexible real-time vision system for automating the detection of potentially dangerous situations at intersections [25]. In this work, vehicles are detected using background subtraction by binarizing the gradient of the moving map, with 90% of successful detection in good conditions. This strategy makes the process more robust to minor shadows or noise, and does not require the choice of an adaptive threshold. However, the system cannot work at night, because lights on the road surface have a major negative impact on the algorithm. To overcome this issues and develop a system that may work over long periods even by night, Arvind [48] proposed the use of infrared cameras (Fig. 16), alongside two corners and in the center of the intersection.

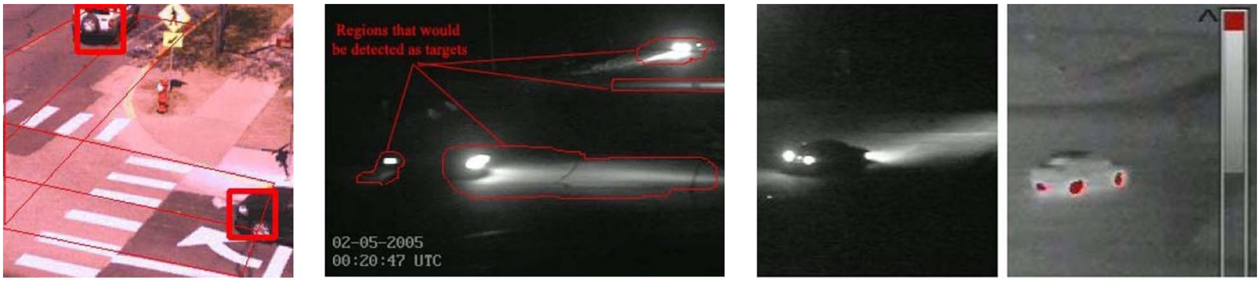
In order to improve the performance of the monitoring at intersections, a multi-camera system has been developed in [29] (Fig. 19). Thanks to the overlapping views, the intrinsic and extrinsic camera calibrations are done by matching vanishing points and lines using visible pattern on the road. Vehicles are segmented based on the Mixture-of-Gaussian (MoG) algorithm for each camera. Then the segmented blobs inputs from each camera are matched, in order to give additional evidence of the vehicle from different angles.

b) ***Omndirectional vision****:* it has been recently introduced for intersection monitoring. A single omnidirectional camera can replace several traditional video-surveillance cameras normally required for monitoring large intersections. However for intersections scene monitoring, a large quantity of pixels is useless and does not cover the active areas of the scene. Therefore, Ghorayeb *et al.* [31] designed an optimized catadioptric vision sensor that increases the useful surface coverage in the image. The mirror-camera needs to be installed nine meters above the center of the crossroad. The performance of the system was demonstrated as they obtained 90% of useful image area, and proceeded to vehicle detection from virtual and real outdoor data [31]. However, despite this performance, the overall system deployment remains bulky and might not be flexible for different intersections.

Fisheye optics has been more often used for intersection monitoring. In [32], the vehicle detection method consists in a background subtraction and a pixel displacement analysis. Jeffery *et al.* [33] proposed a fisheye vision system for data collection purposes about vehicles, motorcycles, bicycles, as well as pedestrians. The vision sensor is required to be pole-mounted at an elevated position in a corner and directed downward the road surface. The system enables to view in a single distorted image all the roads of an intersection. The authors applied background subtraction and model-based verification. Only objects that move along consistent trajectories in appropriate directions are kept active and monitored [33], [74] (Fig. 21). Lately, the research has evolved and led to the development of commercial applications for intersection safety, based on [33], [74], in order to ameliorate the user experience of traffic controllers [34].

2) ***In-vehicle*** Intersection Monitoring Systems:

In-vehicle systems are mainly based on **stereo-vision** setups installed in front of the vehicle. These platforms with in-vehicle cameras have been the focus of research in recent years for Advanced Driver Assistance Systems and traffic modeling. Barth *et al.* [163] developed a stereo-vision in-vehicle system that calculates the optical flow to estimate trajectories of vehicles represented as rigid 3D points clouds. Aycard *et al.*, [39] developed a stereo vision-based demonstrator, for conflict detection at intersections. In this work, cameras are placed in front looking forward in a region up to 35 m in depth, with 70° horizontal field of view (Fig. 22(b)). A two-level architecture stereo



(a) (b) (c)

Fig. 16. (a) Vehicle detection from a visible camera. (b) Effects of headlights in night-time vision. (c) Example use of vehicle detection with infrared cameras by night using the hypothesis of wheel temperatures [48] (2005).

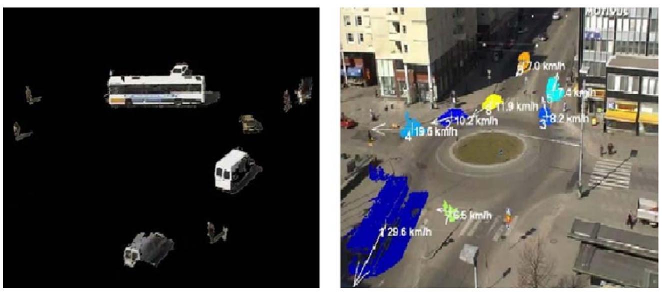


Fig. 17. Vehicle detection and tracking at intersections by 3D-connected component analysis [162] (2005).

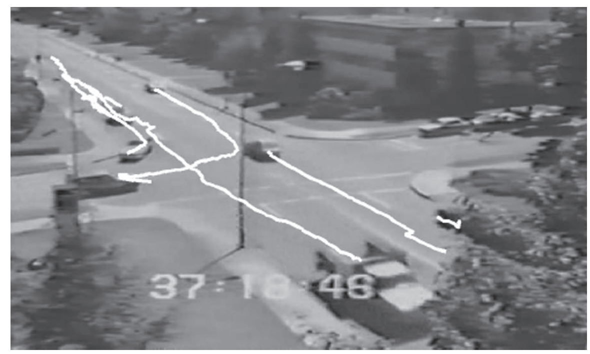


Fig. 18. A feature-based vehicle tracking at intersections [77] (2006).

sensor provides 6D-point information (3D position and 3D motion) using sparse optical flow of corners. Paromtchik *et al.* [164] worked on multimodal vehicle detection, based on data fusion from telemetric sensors and stereo-vision by means of the Bayesian Occupancy Filter (BOF). The authors showed that stereo-vision can enable vehicle detection at up to 10 m and discussed the advantage of sensor fusion. Muffert *et al.* [165] proposed to detect vehicles, represented by dynamic stixels, at a roundabout and warn the driver after time-to-contact computation (Fig. 23). Though several demonstrators have been developed for in-vehicles systems, the technology itself is yet to be transferred and generalized for public use.

C. *Vision-Based Vehicle* ***Tracking at Intersections***

1) ***Roadside*** *Intersection Monitoring Systems:* Many tracking approaches for roadside intersection monitoring are region-based and assume predictable vehicle motions [25]. Feature-based approaches can be used to handle partial occlusions in intersections [66]. However, the difficulties of grouping features (over-grouping), especially in far distance from the camera, can introduce important errors during vehicle tracking at intersections [77] (Fig. 18).

Tracking problems in intersections have been often addressed by the mean of Bayesian frameworks. The Kalman filter was used in [166] for a multiple-target tracking system at crossroad traffic. The proposed mechanism constructs candidates' measurement list by first matching the sizes of the measurements and the targets. It is intended for tracking occluded vehicles without important computational complexity. For each object in the tracking list that has a vehicle ID, Kalman filtering is applied to predict its position in the next frame. This method gives an accuracy higher than 96%.

In [66], vehicle tracking at urban intersections has been tackled by using Spatial-Temporal Markov Random Field, modeled as a graph. The input image is reduced to smaller blocks which represent nodes in the graph. A solution for the object map in the current image, is found based on the previous frames and the previous object map. The result is used in a Hidden Markov Model (HMM) to detect events like vehicle collisions. In [70], graph correspondence was also proposed to track objects segmented by Gaussian Mixture Model. The objects are classified based on the main orientation of their bounding box. Example images for different weather conditions are shown without quantitative results.

In [147], the proposed approach was based on Markov Chain Monte Carlo sampling within a Bayesian framework. First, a foreground map is computed using background subtraction. A proposal map is computed from the foreground map indicating likely vehicle centroids. The distance of points from the boundary of the foreground map indicates the likelihood in the proposal map. A Bayesian problem is formulated for the vehicle positions. The proposal eliminates overlapping vehicles in 3D space and is evaluated by the match between foreground map and projection silhouettes of the 3D models. Tracking between frames is performed by a Viterbi Optimization algorithm which finds the optimal track through the set of solutions for every frame.

In [25], the system uses explicit 3D modeling to track vehicles at intersections. The 3D models are used to initialize an object list for every fifth frame based on the convex hull overlap of model projection and motion map, with the camera calibration information. A feature tracker follows the detected objects along some frames before a new initialization takes place. The tracker is used to speed up operation, as the 3D operation would not be fast enough to operate on every frame in real time. The performance was evaluated on 45 minutes of video from two different sites, with 91.5% reliability of the vehicle counter on test data.

Fig. 19. A multicamera tracking system at intersections [29] (2013).

Fig. 20. Intersection monitoring with large perspective deformation [67] (2014).

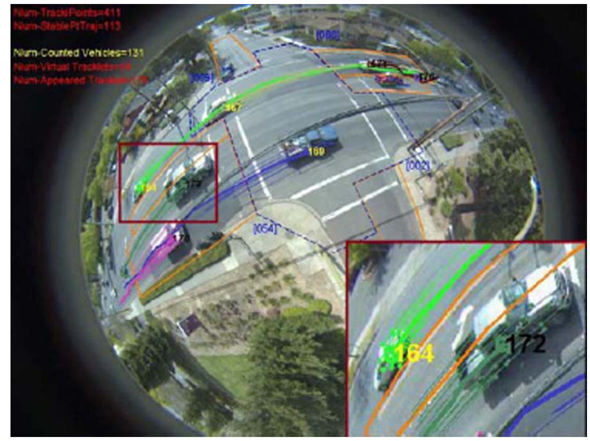
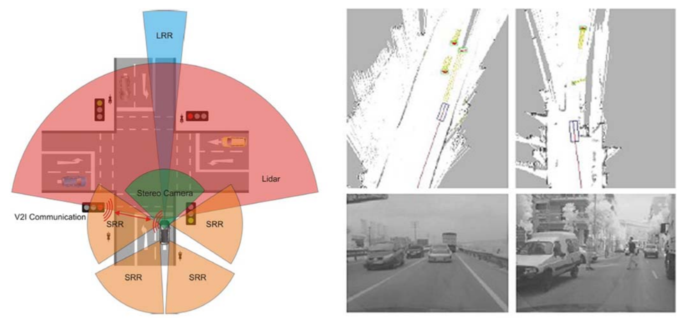


Fig. 21. Real-time multivehicle tracking at intersections from a fisheye camera [74] (2015).



(a) (b)

Fig. 22. Intersection safety platform developed in the project INTERSAFE. (a) Sensor setup. (b) Mapping and moving-object detection results [39] (2011).

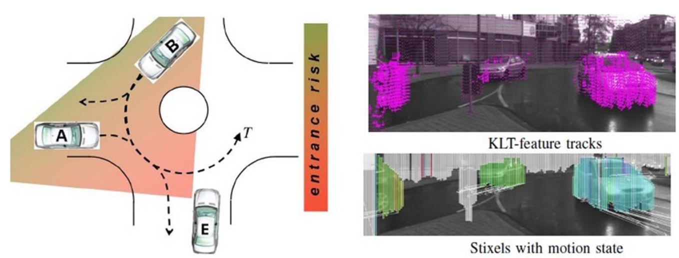


Fig. 23. Vehicle tracking by dynamic stixels and time-to-contact computation (typical situation where the ego-vehicle E has to consider both cars A and B) [165] (2012).

In [77], interest points are tracked independently at urban intersections by the mean of KLT algorithm. This provides robustness against errors in the background estimation and can deal with changing viewing angle, as no prior assumption to the constellation of feature points is made. The tracking performance is reported between 85% and 94%.

*2) **In-Vehicle** Intersection Monitoring Systems:* For in-vehicle systems it appears that Bayesian frameworks are more common. In [37], [163], both motion and depth information are combined to estimate the pose and motion parameters of an oncoming vehicle, including the yaw rate, by means of Kalman filters. In order to cover the dynamic range of a vehicle, an Interacting Multiple Model (IMM) filtering is proposed. It is able to automatically choose the right motion model for typical urban scenarios. Moreover, a gauge consistency criteria, as well as a robust outlier detection method allowing for dealing with sudden accelerations and self-occlusions during turn maneuvers is introduced.

In [39], the authors deal with tracking multiple objects in an intersection like-scenario from a movable platform. A data association step is first performed in order to assign new objects to the existing tracks, and then optimized thanks to information provided by stereo vision. In fact, at an intersection there may be many objects moving in different directions, vehicles may be crossing or waiting to cross in a direction perpendicular to other oncoming vehicles. Tracks are validated or deleted using the outputs from the data association step. In addition, an on-line adapting version of Interacting Multiple Models (IMM) filtering technique is adopted: four Kalman filters are used to handle four motion models (constant velocity, constant acceleration, left turn and right turn). The output of the tracking process consists of: position and velocity information of the ego vehicle along with a list of moving object with their respective position, orientation, velocity and classification information as well as a reference to their instance in the previous frames.

In [37], a generic system for vehicle tracking, in which objects are modeled as rigid 3D point clouds moving along circular paths [165] was proposed. An extended Kalman filter is used for estimating the pose and motion parameters, including velocity, acceleration, and rotational velocity in terms of yaw rate. This feature-based approach also includes geometrical constraints that require the estimated object pose to be consistent with object silhouettes derived from dense stereo data. The Kalman filter allows for modeling a particular expected dynamic behavior of a tracked instance at intersections, where there are typically two options: straight motion or turn movement. The former is best modeled by a stationary process of constant velocity linear motion, while turning vehicles require higher-order motion models.

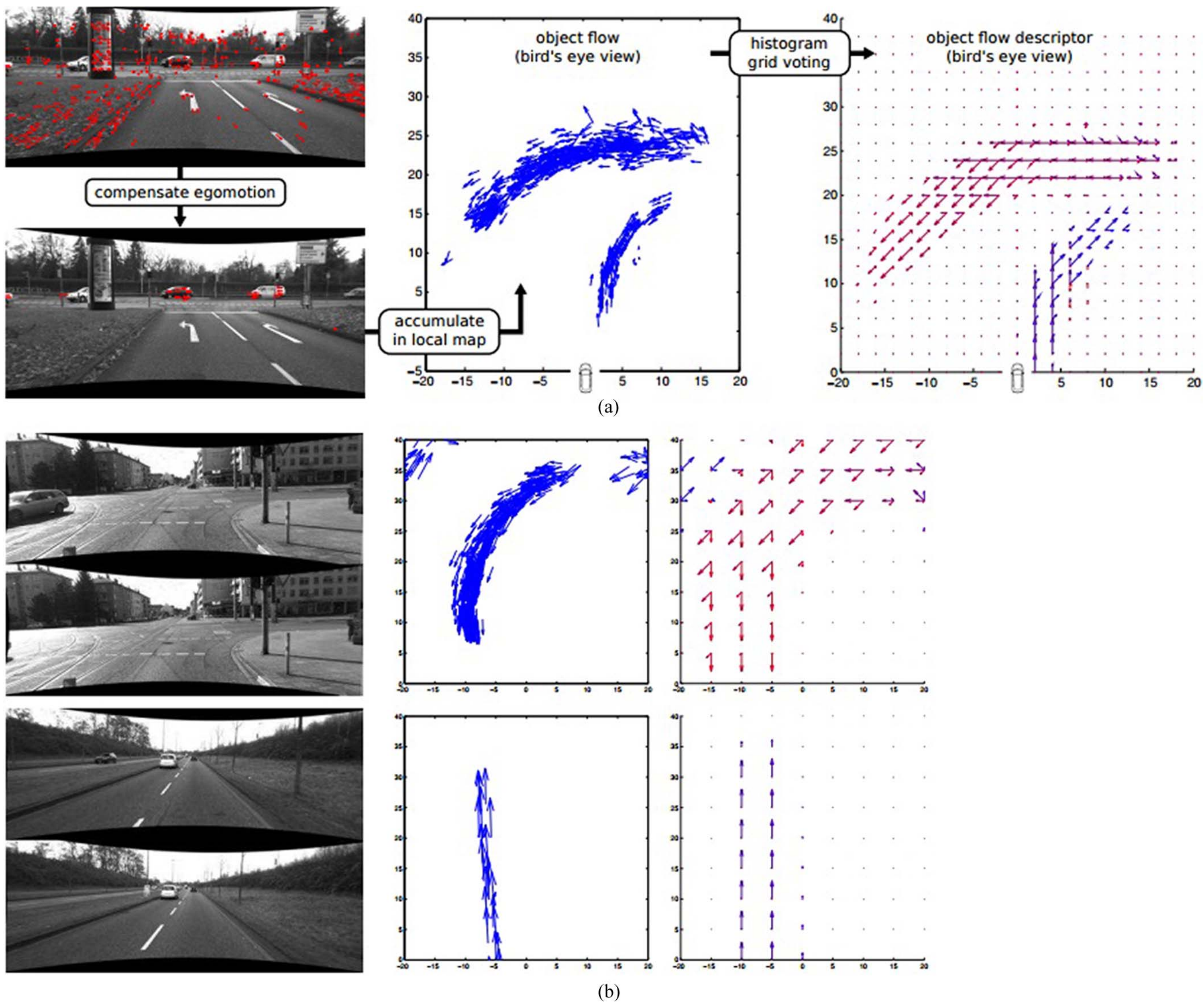


Fig. 24. Vision-based infrastructure detection with in-vehicle systems: detect when vehicles reach an intersection [121] (2011). (a) A stereo-based motion descriptor that registers flow vectors overtime by visual odometry. Then, votes cast by flow vectors are accumulated into a histogram to detect when the vehicle reaches an intersection. (b) Examples of detections: intersection and nonintersection.

D. Vehicle Behavior Analysis at Intersections

Some intersection monitoring systems are intended to count traffic participants classified in different categories [74]. However, the goal of most intersection monitoring systems is to analyze the interactions between vehicles at intersections in order to detect abnormal events and prevent accidents. Generally a potential accident or collision is defined when "unless the speed and/or the direction of the road users changes, they will collide" [167], [168]. Thus, a robust event detection algorithm must be independent of geometric factors, such as geometry of the intersection, angle of video camera, and position where the accident occurred [66].

A roadside system which relies on two databases was presented in [169]. First, a trajectory database, where the results of the vehicle tracking module are stored, with the generation of a distribution over possible future positions given previous positions for each road user at any instant. Second, an interaction database is created, where trajectories relations between road users are considered with several indicators. Besides, the knowledge about regular road user motion patterns are used to reduce the possibilities, in order to propose more realistic and accurate motion prediction (Fig. 25).

For in-vehicles systems, trajectory prediction can be as important as scene understanding [170]. In [21], a conditionally independent probabilistic model that integrates a fusion of multiple cues was proposed (Fig. 26). On the one hand it allows autonomous vehicles to estimate the layout of urban intersections based on in-vehicle stereo vision. On the other hand, vehicle motions are learned using maximum likelihood and contrastive divergence in order to infer driving directions. The estimation of time-to-contact (TTC) is a good indicator of a potential conflict, and generally compared with the commonly used total reaction time of 2 seconds. In [165], the goal was to predict whether a safe entrance into the roundabout is possible or not, in front of the ego-vehicle. A least square mechanism is used to fit a cluster of vehicle trajectories over time. The TTC is obtained as the ratio of the length of the circular fitted arc by the

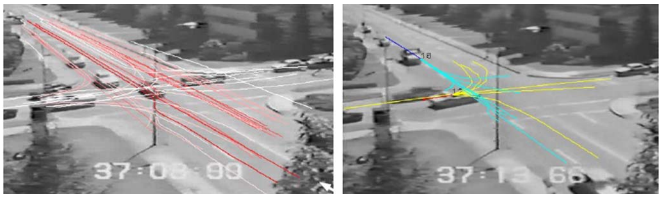


Fig. 25. Probabilistic collision prediction for roadside intersection monitoring. (Left) Training of traffic conflicts using prototype trajectories. (Right) An example of movement prediction; the vehicle trajectories are red and blue, with a dot marking their position, and the future positions are respectively cyan and yellow [169].

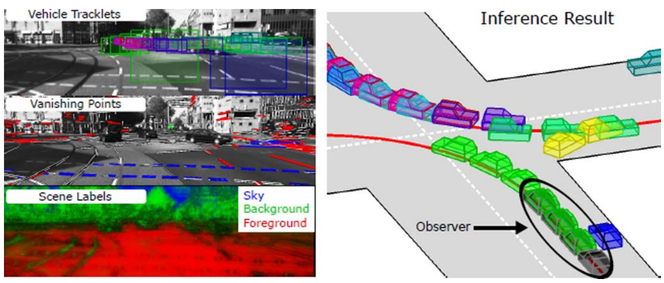


Fig. 26. Use of appearance and motion feature cues to infer the road layout and the location of traffic participants in the scene from short video sequences [21].

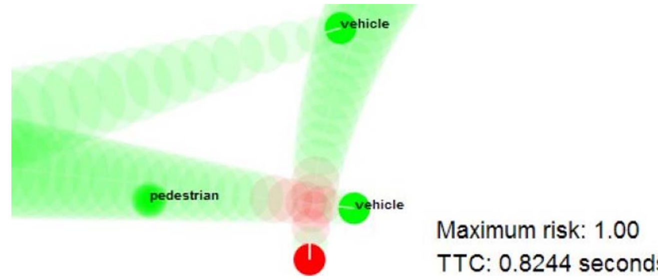


Fig. 27. Example of a potential collision between the host vehicle (red circle on the bottom) and another vehicle (green circle on the bottom right) [39].

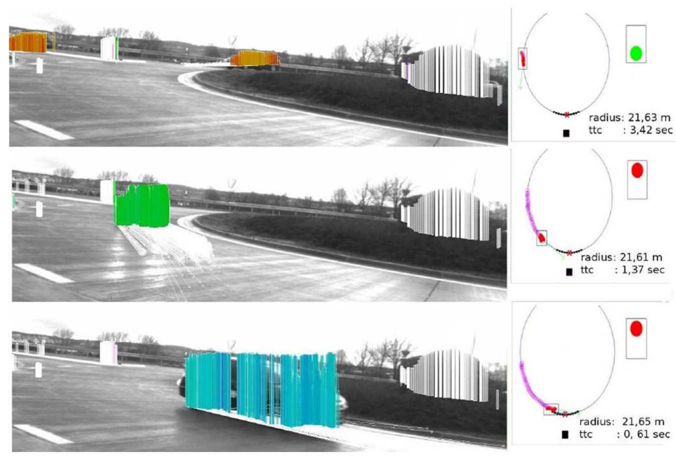


Fig. 28. Extract of a sequence: typical scene with an incoming vehicle at a roundabout. On the right side: The results of the clustering process and the TTC computation are shown for each scene [165].

mean velocity (Fig. 28). In [39], a dynamic circle-based strategy has been used for frontal and lateral collision prediction. In this project, the host vehicle and dynamic objects are represented as circles. A potential collision is detected when the host vehicle circle intersects at least one circle of the dynamic objects at the same time (Fig. 27).

V. Summary and Discussion of Future Trends of Traffic Monitoring at Intersections

The current state-of-the art of vehicle detection and tracking at intersections can be broken into two categories: roadside and in-vehicle systems. There has been several advances in both areas and many papers have reported excellent statistics, however the monitoring performance is still affected in many cases by the environment (illumination, shadows, night time), scene structure (large baseline, different intersections types) and particularly by the vehicle motions at intersections (occlusion, change in appearance, abrupt motions).

A. Roadside Systems vs In-Vehicle Systems

1) System Setup: Roadside systems consist generally in pole-mounted cameras fixed on masts, sometimes in the corner or in the center of the intersections. Most of existing systems are not designed to be flexible to different types of intersections and do not offer a wide field of view at the intersection. In the recent literature, omnidirectional cameras have been used. Furthermore, in most studies related to roadside intersection monitoring, a user is required to execute several preprocessing tasks. Therefore, the accuracy of the detection methods highly depends on the parameters set by the user. A challenge would be to develop automatic calibration methods and a general pipeline for traffic monitoring at intersection that would require the least human intervention. Roadside monitoring systems are intended to work for many hours, therefore data storage and transmission must be optimized to fulfill this requirements [25], [38].

In-vehicle systems for intersection monitoring are developed for Advanced driver Assistance Systems (ADAS). Vision-based systems in this category use stereo setups in order to monitor traffic in front of the vehicle when it reaches the intersection. Cooperative sensor fusion with Lidar and Radar, has been commonly used in this area lately. Also, these systems are generally intended for research and are hardly immediately available for general public use.

2) Vehicle Detection and Tracking Methods: For roadside applications, most real-world vehicle detection systems are based on variations of background subtraction algorithms. This is because methods such as model-based detection have a high computational complexity and barely meets the real time requirements [74]. However, background subtraction can result in important segmentation errors for cluttered scenes and occluded vehicles. Therefore, in the recent literature of roadside systems, vehicle tracking are mainly based on sparse feature matching with the popular KLT algorithm. Additional motion, model cues and prior knowledge of the scene are also necessary to accurately group features and correctly classify traffic participants.

For in-vehicle applications, the first objective is the detection and tracking of upcoming or turning frontal vehicles. Motion cues have been mostly used with different adaptations of the optical flow. Once vehicle are detected, they are generally tracked using generic Bayesian algorithms, especially Kalman and particle filters. The optical flow has also been used to detect when vehicles arrive at an intersection (Fig. 24).

3) Worthiness of Omnidirectional Vision for Intersection Monitoring: omnidirectional vision allows to monitor an entire intersection, but introduces several challenges because of the large amount of visual information obtained in a single image. There are few omnidirectional vision-based systems for intersections monitoring. An optimized catadioptric camera and fish-eye optics have been used within the recent literature.

A network of calibrated omnidirectional cameras could be used to obtain rich 3D/4D reconstruction of intersections; which has not been done in any related works. This would allow to improve the accuracy of vehicle monitoring at intersections and can be generalized to any type of intersections. It would also provide an excellent realistic augmented map of the traffic at the intersection. Finally, the system would also need to be flexible to different types of intersections, robust to difficult weather, without requiring important civil engineering, and regular maintenance. That is to say that the design of a real-time omnidirectional-vision based vehicle detection and tracking system, flexible to intersections geometries and easily reconfigurable, remains a research subject worth exploring for intelligent transportation systems.

VI. Conclusion and Perspectives

In this paper, we have presented a review of vision-based vehicle detection and tracking systems for intersection monitoring. There has been an increase of studies focused on intersections over the last decade, with both in-vehicle and roadside systems. The former are often based on stereo-vision, whereas the latter use generally monocular vision and more recently pole-mounted omnidirectional cameras. Omnidirectional cameras provide a wide field of view and can be used to monitor all the roads of an intersection in a single image. However there are several computer vision challenges that arise when using omnidirectional cameras at intersections. In particular, it is necessary to process a large amount of visual information in real time and to monitor objects that have variable scales in the same image and can have different motion patterns. At the same time, it is also necessary to take care of classical vehicle detection and tracking issues caused by occlusion, presence of shadows, abrupt changes in the environment or the scene. There are major possible contributions directions, especially for stationary systems: the automation of the initialization steps such as camera calibration by directly exploiting vehicles motions; the development of a framework for general fusion of active and passive sensors for night time vision and difficult weather; the development of a plug-and-play tools for fast diagnosis. Besides, the use of multi-cameras systems has been hardly investigated for intersections, certainly due to the large baselines. Multiple omnidirectional cameras could be used in a calibrated network, to facilitate the detection of occluded vehicles and improve their evidence from several views. Moreover multiple-view geometry could be studied to improve the accuracy of the monitoring, using 3D and 4D reconstruction at the intersection.

Acknowledgment

This work was initiated within the framework of the project GRR LMN ROADTRAC (ROAD TRaffic Analysis by Crowdsourcing) and carried out within the Ph.D. project "Traffic Diagnosis of Road Intersections by Omnidirectional Vision."

References

[1] M. Williams, "The Prometheus programme," in *Proc. IEE Colloq. Towards Safer Road Transp.—Eng. Sol.*, Mar. 1992, pp. 4/1–4/3.

[2] B. Ulmer, "VITA—An Autonomous Road Vehicle (ARV) for collision avoidance in traffic," in *Proc. Intell. Veh. Symp.*, Jun. 1992, pp. 36–41.

[3] R. T. Collins *et al.*, "A system for video surveillance and monitoring," Carnegie Mellon Univ., Robotics Inst., Pittsburg, PA, USA, 2000, vol. 2.

[4] M. Naylor and C. I. Attwood, "Annotated digital video for intelligent surveillance and optimized retrieval: Final report," ADVISOR Connortium, 2003.

[5] 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.

[6] B. T. Morris and M. M. Trivedi, "Understanding vehicular traffic behavior from video: A survey of unsupervised approaches," *J. Electron. Imag.*, vol. 22, no. 4, 2013, Art. ID 041113.

[7] L. A. Rodegerdts, *Roundabouts: An Informational Guide*, vol. 672. Washington, DC, USA: Transportation Research Board, 2010.

[8] S. Lefèvre, C. Laugier, and J. Ibañez-Guzmán, "Risk assessment at road intersections: Comparing intention and expectation," in *Proc. IEEE IV Symp.*, 2012, pp. 165–171.

[9] Federal Highway Administration, "The national intersection safety problem," U.S. Dept. Transp., FHWA-SA-10-005, Nov. 2009. Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/kvLIYI

[10] S. Hoeglinger *et al.*, Traffic Safety Basic Facts 2007: Junctions, 2008. Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/AE8SYX

[11] G. Yannis, C. Antoniou, and P. Evgenikos, "Comparative analysis of junction safety in Europe," in *Proc. 12th WCTR*, 2010, pp. 1–11.

[12] European Commission, Traffic Safety Basic Facts 2012: Junctions, 2013. Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/YYpqD6

[13] European Commission, Traffic Safety Basic Facts 2015: Junctions, Jun. 2015. Accessed on: Dec. 6, 2015. [Online]. Available: http://goo.gl/ZkhX15

[14] P. Subirats, Y. Dupuis, E. Violette, D. Doucet, and G. Dupre, "A new tool to evaluate safety of crossroad," in *Proc. 4th Int. Symp. Highway Geometric Des.*, Valencia, Spain, 2010, pp. 2–5.

[15] European Commission, Road Safety in the European Union: Trends, Statistics and Challenges, Mar. 2015. Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/LqWwl7

[16] 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.

[17] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image Vis. Comput.*, vol. 21, no. 4, pp. 359–381, Apr. 2003.

[18] H. Guo and Z. Wang, "TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection," in *Proc. IEEE PacificVis Symp.*, 2011, pp. 163–170.

[19] E. Kafer, C. Hermes, C. Wohler, H. Ritter, and F. Kummert, "Recognition of situation classes at road intersections," in *Proc. IEEE ICRA*, 2010, pp. 3960–3965.

[20] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2287–2301, Nov. 2011.

[21] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3D traffic scene understanding from movable platforms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 1012–1025, May 2014.

[22] H. Zhang, A. Geiger, and R. Urtasun, "Understanding high-level semantics by modeling traffic patterns," in *Proc. IEEE ICCV*, 2013, pp. 3056–3063.

[23] E. Dabbour and S. M. Easa, "Perceptual framework for a modern left-turn collision warning system," *Int. J. Civil Environ. Struct. Construct. Architect. Eng.*, vol. 33, no. 9, pp. 640–646, 2009.

[24] N. Saunier and T. Sayed, "Automated analysis of road safety with video data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2019, no. 1, pp. 57–64, 2007.

[25] S. Messelodi, C. 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.

[26] B. Rinner and W. Wolf, "An introduction to distributed smart cameras," *Proc. IEEE*, vol. 96, no. 10, pp. 1565–1575, Oct. 2008.

[27] Y. Shi and F. D. Real, "Smart cameras: Fundamentals and classification," in *Smart Cameras*. New York, NY, USA: Springer-Verlag, 2010, pp. 19–34.

[28] Z. Hu, C. Wang, and K. Uchimura, “3D vehicle extraction and tracking from multiple viewpoints for traffic monitoring by using probability fusion map,” in *Proc. IEEE ITSC*, 2007, pp. 30–35.

[29] H. Tang, “Development of a multiple-camera tracking system for accurate traffic performance measurements at intersections,” Intell. Transp. Syst. Inst., Center Transp. Studies, Univ. Minnesota Duluth, Duluth, MN, USA, CTS 13-10, 2013.

[30] M. M. Trivedi, T. Gandhi, and J. McCall, “Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety,” *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 108–120, Mar. 2007.

[31] A. Ghorayeb *et al.*, “ Capteur omnidirectionnel optimal pour le diagnostic de la circulation dans les carrefours urbains,” presented at the ORASIS–Congrès des Jeunes Chercheurs en Vision par Ordinateur, Trégastel, France, 2009.

[32] S.-M. Lee and H. Baik, “Origin–Destination (OD) trip table estimation using traffic movement counts from vehicle tracking system at intersection,” in *Proc. 32nd Annu. Conf. IEEE IECON*, 2006, pp. 3332–3337.

[33] T. F. Gee and J. R. Price, “Omnidirectional imaging and computer vision for transportation applications: From conception to deployment,” AldisCorp, Knoxville, TN, USA, Report, 2011.

[34] GRIDSMART, Traffic Management Systems for Intersection Monitoring. Accessed on: Oct. 1, 2015. [Online]. Available: https://gridsmart.com

[35] Intelligent Transportation Systems, Tokyo, Japan, Green Safety: Public-Private Collaborative Projects That Aims “for a Greener and Safer Society” Using Cooperative Systems. Accessed on: Oct. 1, 2015. [Online]. Available: http://www.its-jp.org/english/its-green-safety-showcase/

[36] S. Gehrig, C. Rabe, and L. Krüger, “6D vision goes fisheye for intersection assistance,” in *Proc. Can. Conf. CRV*, 2008, pp. 34–41.

[37] D. Pfeiffer, A. Barth, and U. Franke, “Vehicle tracking at urban intersections using dense stereo,” in *Proc. 3rd Workshop BMI*, Ghent, Belgium, Nov. 2009, pp. 47–58.

[38] H. Zhao *et al.*, “Sensing an intersection using a network of laser scanners and video cameras,” *IEEE Intell. Transp. Syst. Mag.*, vol. 1, no. 2, pp. 31–37, Summer 2009.

[39] O. Aycard *et al.*, “Intersection safety using lidar and stereo vision sensors,” in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 863–869.

[40] M. Goldhammer *et al.*, “Cooperative multi sensor network for traffic safety applications at intersections,” in *Proc. 15th IEEE ITSC*, 2012, pp. 1178–1183.

[41] E. Strigel, D. Meissner, F. Seeliger, B. Wilking, and K. Dietmayer, “The Ko-PER intersection laser scanner and video dataset,” in *Proc. IEEE 17th ITSC*, 2014, pp. 1900–1901.

[42] Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection: A review,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.

[43] 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.

[44] S. Sivaraman and M. M. Trivedi, “Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking and behavior analysis,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[45] L. Alexander *et al.*, “The Minnesota mobile intersection surveillance system,” in *Proc. IEEE ITSC*, 2006, pp. 139–144.

[46] J. Sha *et al.*, “Trajectory analysis of moving objects at intersection based on laser-data,” in *Proc. 14th IEEE ITSC*, 2011, pp. 289–294.

[47] Q. Zhu *et al.*, “3D lidar point cloud based intersection recognition for autonomous driving,” in *Proc. IEEE IV Symp.*, 2012, pp. 456–461.

[48] M. Arvind, “Sensor fusion for real-time gap tracking and vehicle trajectory estimation at rural intersections,” Ph.D. dissertation, Univ. Minnesota, Minnesota, MN, USA, 2005.

[49] R. J. Radke, “A survey of distributed computer vision algorithms,” in *Handbook of Ambient Intelligence and Smart Environments*. New York, NY, USA: Springer-Verlag, 2010, pp. 35–55.

[50] M. S. Shirazi and B. Morris, “Contextual combination of appearance and motion for intersection videos with vehicles and pedestrians,” in *Advances in Visual Computing*. Cham, Switzerland: Springer-Verlag, 2014, pp. 708–717.

[51] X. Wang, X. Ma, and W. E. L. Grimson, “Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 539–555, Mar. 2009.

[52] U.S. Dept. Transp., Next Generation Simulation Systems (NGSIM). Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/zHOX5M

[53] C. Thiemann, M. Treiber, and A. Kesting, “Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data,” *Transp. Res. Rec., J. Transp. Res. Board*, no. 2088, pp. 90–101, 2008.

[54] Z. Sun and X. J. Ban, “Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors,” *Transp. Res. C, Emerging Technol.*, vol. 36, pp. 268–283, Nov. 2013.

[55] P. Hao, X. Ban, K. P. Bennett, Q. Ji, and Z. Sun, “Signal timing estimation using sample intersection travel times,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 792–804, Jun. 2012.

[56] X. J. Ban and M. Gruteser, “Towards fine-grained urban traffic knowledge extraction using mobile sensing,” in *Proc. ACM SIGKDD Int. Workshop Urban Comput.*, 2012, pp. 111–117.

[57] Ö. Aköz and M. E. Karsligil, “Traffic event classification at intersections based on the severity of abnormality,” *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 613–632, Apr. 2014.

[58] W. Hu *et al.*, “A system for learning statistical motion patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, Sep. 2006.

[59] C. C. Loy, T. Xiang, and S. Gong, “Detecting and discriminating behavioural anomalies,” *Pattern Recognit.*, vol. 44, no. 1, pp. 117–132, Jan. 2011.

[60] T. Hospedales, S. Gong, and T. Xiang, “Video behaviour mining using a dynamic topic model,” *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 303–323, Jul. 2012.

[61] KIT Intersection Monitoring Datasets in Various Wheather. Accessed on: Oct. 1, 2015. [Online]. Available: http://goo.gl/wLlnIN

[62] 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.

[63] J.-P. Jodoin, G.-A. Bilodeau, and N. Saunier, “Urban tracker: Multiple object tracking in urban mixed traffic,” in *Proc. IEEE WACV*, 2014, pp. 885–892.

[64] U. Muehlmann, M. Ribo, P. Lang, and A. Pinz, “A new high speed CMOS camera for real-time tracking applications,” in *Proc. IEEE ICRA*, 2004, vol. 5, pp. 5195–5200.

[65] P. Babaei, “Vehicles tracking and classification using traffic zones in a hybrid scheme for intersection traffic management by smart cameras,” in *Proc. ICSIP*, 2010, pp. 49–53.

[66] 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.

[67] T. Furuya and C. J. Taylor, “Road intersection monitoring from video with large perspective deformation,” in *Proc. 21st World Congr. Intell. Transp. Syst.*, Sep. 2014, pp. 1–12.

[68] K. Ismail, T. Sayed, and N. Saunier, “Camera calibration for urban traffic scenes: Practical issues and a robust approach,” in *Proc. Transp. Res. Board Annu. Meet. Compendium Papers*, 2010, pp. 1–20.

[69] D. S. Ly, C. Demonceaux, P. Vasseur, and C. Pégard, “Extrinsic calibration of heterogeneous cameras by line images,” *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1601–1614, Aug. 2014.

[70] H. Vceraraghavan, O. Masoud, and N. Papanikolopoulos, “Vision-based monitoring of intersections,” in *Proc. IEEE 5th Int. Conf. Intell. Transp. Syst.*, 2002, pp. 7–12.

[71] N. K. Kanhere and S. T. Birchfield, “A taxonomy and analysis of camera calibration methods for traffic monitoring applications,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 441–452, Jun. 2010.

[72] M. Dubska, J. Sochor, and A. Herout, “Automatic camera calibration for traffic understanding,” in *Proc. BMVC*, 2014, pp. 1–10.

[73] H. Veeraraghavan, O. Masoud, and N. P. Papanikolopoulos, “Computer vision algorithms for intersection monitoring,” *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 2, pp. 78–89, Jun. 2003.

[74] W. Wang, T. Gee, J. Price, and H. Qi, “Real time multi-vehicle tracking and counting at intersections from a fisheye camera,” in *Proc. IEEE WACV*, 2015, pp. 17–24.

[75] N. Buch, J. Orwell, and S. A. Velastin, “Detection and classification of vehicles for urban traffic scenes,” in *Proc. 5th Int. Vis. Inf. Eng.*, 2008, pp. 182–187.

[76] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.

[77] N. Saunier and T. Sayed, “A feature-based tracking algorithm for vehicles in intersections,” in *Proc. 3rd Can. Conf. Comput. Robot Vis.*, Jun. 2006, pp. 59.

[78] O. Barnich and M. Van Droogenbroeck, “ViBe: A universal background subtraction algorithm for video sequences,” *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[79] A. Sobral and A. Vacavant, “A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos,” *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[80] A. Sanin, C. Sanderson, and B. C. Lovell, “Shadow detection: A survey and comparative evaluation of recent methods,” *Pattern Recognit.*, vol. 45, no. 4, pp. 1684–1695, Apr. 2012.

[81] T. Bucher *et al.*, "Image processing and behavior planning for intelligent vehicles," *IEEE Trans. Ind. Electron.*, vol. 50, no. 1, pp. 62–75, Feb. 2003.

[82] Y. Zhu, D. Comaniciu, M. Pellkofer, and T. Koehler, "Reliable detection of overtaking vehicles using robust information fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 401–414, Dec. 2006.

[83] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through 'v-disparity' representation," in *Proc. IEEE Intell. Veh. Symp.*, 2002, vol. 2, pp. 646–651.

[84] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and tracking the driving environment with a particle-based occupancy grid," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1331–1342, Dec. 2011.

[85] Y.-C. Lim, M. Lee, C.-H. Lee, S. Kwon, and J.-H. Lee, "Improvement of stereo vision-based position and velocity estimation and tracking using a stripe-based disparity estimation and inverse perspective map-based extended Kalman filter," *Opt. Lasers Eng.*, vol. 48, no. 9, pp. 859–868, Sep. 2010.

[86] J. M. Wang *et al.*, "Vision-based traffic measurement system," in *Proc. 17th ICPR*, 2004, vol. 4, pp. 360–363.

[87] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino, "Background subtraction for automated multisensor surveillance: A comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, Aug. 2010, Art. ID 343057.

[88] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Vis. Pattern Recog.*, 1999, vol. 2, pp. 246–252.

[89] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of Gaussians for foreground detection—A survey," *Recent Patents Comput. Sci.*, vol. 1, no. 3, pp. 219–237, Nov. 2008.

[90] Y. Guo *et al.*, "Matching vehicles under large pose transformations using approximate 3D models and piecewise MRF model," in *Proc. IEEE CVPR*, 2008, pp. 1–8.

[91] W. Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, "Traffic accident prediction using 3-D model-based vehicle tracking," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 677–694, May 2004.

[92] J. Lou, T. Tan, W. Hu, H. Yang, and S. J. Maybank, "3-D model-based vehicle tracking," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1561–1569, Oct. 2005.

[93] 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.

[94] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.

[95] G. Zhang, R. P. Avery, and Y. Wang, "Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1993, no. 1, pp. 138–147, 2007.

[96] 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.

[97] 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.

[98] J. Nuevo, I. Parra, J. Sjoberg, and L. M. Bergasa, "Estimating surrounding vehicles' pose using computer vision," in *Proc. 13th IEEE ITSC*, 2010, pp. 1863–1868.

[99] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 583–595, Jun. 2011.

[100] C. Hoffmann, "Fusing multiple 2D visual features for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 406–411.

[101] N. Blanc, B. Steux, and T. Hinz, "Larasidecam: A fast and robust vision-based blindspot detection system," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 480–485.

[102] Y.-M. Chan, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Vehicle detection under various lighting conditions by incorporating particle filter," in *Proc. IEEE ITSC*, 2007, pp. 534–539.

[103] J. Arróspide and L. Salgado, "A study of feature combination for vehicle detection based on image processing," *Sci. World J.*, vol. 2014, 2014, Art. ID 196251.

[104] X. Zhang, N. Zheng, Y. He, and F. Wang, "Vehicle detection using an extended hidden random field model," in *Proc. 14th IEEE ITSC*, 2011, pp. 1555–1559.

[105] C.-Y. Hsu, L.-W. Kang, and H.-Y. M. Liao, "Cross-camera vehicle tracking via affine invariant object matching for video forensics applications," in *Proc. IEEE ICME*, 2013, pp. 1–6.

[106] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2019–2034, Jul. 2006.

[107] P. E. Rybski, D. Huber, D. D. Morris, and R. Hoffman, "Visual classification of coarse vehicle orientation using histogram of oriented gradients features," in *Proc.IEEE IV Symp.*, 2010, pp. 921–928.

[108] M. Cheon, W. Lee, C. Yoon, and M. Park, "Vision-based vehicle detection system with consideration of the detecting location," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1243–1252, Sep. 2012.

[109] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 125–137, Jun. 2005.

[110] L. Mao, M. Xie, Y. Huang, and Y. Zhang, "Preceding vehicle detection using histograms of oriented gradients," in *Proc. ICCCAS*, 2010, pp. 354–358.

[111] W. Liu, X. Wen, B. Duan, H. Yuan, and N. Wang, "Rear vehicle detection and tracking for lane change assist," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 252–257.

[112] Z. Yin and R. Collins, "Belief propagation in a 3D spatio-temporal MRF for moving object detection," in *Proc. IEEE CVPR*, 2007, pp. 1–8.

[113] J.-Y. Bouguet, *Pyramidal Implementation of the Affine Lucas Kanade Feature Tracker Description of the Algorithm*, vol. 5. Santa Clara, CA, USA: Intel Corporation, 2001, pp. 1–10.

[114] T. Kowsari, S. S. Beauchemin, and J. Cho, "Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost," in *Proc. 14th IEEE ITSC*, 2011, pp. 1255–1260.

[115] Y. Cao, A. Renfrew, and P. Cook, "Vehicle motion analysis based on a monocular vision system," in *Proc. IET RTIC ITS U.K. Members Conf.*, May 2008, pp. 1–6.

[116] Y. Cao, A. Renfrew, and P. Cook, "Novel optical flow optimization using pulse-coupled neural network and smallest univalue segment assimilating nucleus," in *Proc. Int. Symp. ISPACS*, Nov. 2007, pp. 264–267.

[117] I. Sato, C. Yamano, and H. Yanagawa, Crossing obstacle detection with a vehicle-mounted camera," in *Proc. IEEE IV Symp.*, 2011, pp. 60–65.

[118] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world-a compact medium level representation of the 3D-world," in *Pattern Recognition*. Berlin, Germany: Springer-Verlag, 2009, pp. 51–60.

[119] F. Erbs, A. Barth, and U. Franke, "Moving vehicle detection by optimal segmentation of the dynamic stixel world," in *Proc. IEEE IV Symp.*, 2011, pp. 951–956.

[120] C. Rabe, U. Franke, and S. Gehrig, "Fast detection of moving objects in complex scenarios," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 398–403.

[121] A. Geiger and B. Kitt, "Object flow: A descriptor for classifying traffic motion," in *Proc. IEEE IV Symp.*, 2010, pp. 287–293.

[122] X. Wen, H. Zhao, N. Wang, and H. Yuan, "A rear-vehicle detection system for static images based on monocular vision," in *Proc. 9th ICARCV*, 2006, pp. 1–4.

[123] M. Lin and X. Xu, "Multiple vehicle visual tracking from a moving vehicle," in *Proc.6th ISDA*, 2006, vol. 2, pp. 373–378.

[124] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 2000.

[125] T. Kohonen, "An introduction to neural computing," *Neural Netw.*, vol. 1, no. 1, pp. 3–16, 1988.

[126] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*. Berlin, Germany: Springer-Verlag, 1995, pp. 23–37.

[127] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," in *Proc. 14th Int. Conf. DSP*, 2002, vol. 2, pp. 1019–1022.

[128] H. Cheng, N. Zheng, and C. Sun, "Boosted Gabor features applied to vehicle detection," in *Proc. 18th ICPR*, 2006, vol. 1, pp. 662–666.

[129] Y. Zhang, S. J. Kiselewich, and W. A. Bauson, "Legendre and Gabor moments for vehicle recognition in forward collision warning," in *Proc. IEEE ITSC*, 2006, pp. 1185–1190.

[130] Q. B. Truong and B. R. Lee, "Vehicle detection algorithm using hypothesis generation and verification," in *Emerging Intelligent Computing Technology and Applications*. Berlin, Germany: Springer-Verlag, 2009, pp. 534–543.

[131] 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.

[132] P. Y. Shinzato, V. Grassi, F. S. Osório, and D. F. Wolf, "Fast visual road recognition and horizon detection using multiple artificial neural networks," in *Proc. IEEE IV Symp.*, 2012, pp. 1090–1095.

[133] K. Deb, I. Khan, A. Saha, and K.-H. Jo, "An efficient method of vehicle license plate recognition based on sliding concentric windows and artificial neural network," *Procedia Technol.*, vol. 4, pp. 812–819, 2012.

[134] B.-H. Chen and S.-C. Huang, "Probabilistic neural networks based moving vehicles extraction algorithm for intelligent traffic surveillance systems," *Inf. Sci.*, vol. 299, pp. 283–295, Apr. 2015.

[135] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[136] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE CVPR*, 2001, vol. 1, pp. I-511–I-518.

[137] M. Stojmenovic, "Real time machine learning based car detection in images with fast training," *Mach. Vis. Appl.*, vol. 17, no. 3, pp. 163–172, Aug. 2006.

[138] C. Wu, L. Duan, J. Miao, F. Fang, and X. Wang, "Detection of front-view vehicle with occlusions using AdaBoost," in *Proc. ICIECS*, 2009, pp. 1–4.

[139] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: A comparative study," *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 599–611, Apr. 2014.

[140] Z. Zivkovic, A. T. Cemgil, and B. Krose, "Approximate Bayesian methods for kernel-based object tracking," *Comput. Vis. Image Understand.*, vol. 113, no. 6, pp. 743–749, Jun. 2009.

[141] R. Danescu, S. Nedevschi, M. M. Meinecke, and T. Graf, "Stereovision based vehicle tracking in urban traffic environments," in *Proc. IEEE ITSC*, 2007, pp. 400–404.

[142] A. Barth, "Vehicle tracking and motion estimation based on stereo vision sequences," Ph.D. dissertation, Inst. Geodäsie Geoinf., Univ. Bonn, Bonn, Germany, 2010.

[143] E. Maggio and A. Cavallaro, *Video Tracking: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2011.

[144] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision," in *Proc. IJCAI*, 1981, vol. 81, pp. 674–679.

[145] C. Tomasi and T. Kanade, "Detection and tracking of point features," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, 1991.

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

[147] X. Song and R. Nevatia, "Detection and tracking of moving vehicles in crowded scenes," in *Proc. IEEE Workshop WMVC*, 2007, pp. 1–8.

[148] B. Morris and M. Trivedi, "Robust classification and tracking of vehicles in traffic video streams," in *Proc. IEEE ITSC*, 2006, pp. 1078–1083.

[149] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool, "Coupled object detection and tracking from static cameras and moving vehicles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1683–1698, Oct. 2008.

[150] M. S. Grewal, *Kalman Filtering*. Berlin, Germany: Springer-Verlag, 2011.

[151] X. Cao, J. Lan, P. Yan, and X. Li, "Vehicle detection and tracking in airborne videos by multi-motion layer analysis," *Mach. Vis. Appl.*, vol. 23, no. 5, pp. 921–935, Sep. 2012.

[152] M. Isard and A. Blake, "Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, Aug. 1998.

[153] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, vol. 12. Oxford, U.K.: Oxford Univ. Press, 2009, pp. 656–704.

[154] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[155] Y.-M. Chan, S.-S. Huang, L.-C. Fu, P.-Y. Hsiao, and M.-F. Lo, "Vehicle detection and tracking under various lighting conditions using a particle filter," *IET Intell. Transp. Syst.*, vol. 6, no. 1, pp. 1–8, Mar. 2012.

[156] H. T. Niknejad, A. Takeuchi, S. Mita, and D. McAllester, "On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 748–758, Jun. 2012.

[157] F. Bardet and T. Chateau, "MCMC particle filter for real-time visual tracking of vehicles," in *Proc. 11th IEEE ITSC*, 2008, pp. 539–544.

[158] A. Ghorayeb, A. Potelle, L. Devendeville, and E. M. Mouaddib, "Optimal omnidirectional sensor for urban traffic diagnosis in crossroads," in *Proc. IEEE IV Symp.*, 2010, pp. 597–602.

[159] H. Dinh and H. Tang, "Simple method for camera calibration of roundabout traffic scenes using a single circle," *IET Intell. Transpo. Syst.*, vol. 8, no. 3, pp. 175–182, May 2013.

[160] H. Tang and H. Dinh, "A tracking-based traffic performance measurement system for roundabouts and intersections," Intell. Transp. Syst. Inst., Center Transp. Studies, Univ. Minnesota, Minnesota, MN, USA, 2012.

[161] H. Veeraraghavan and N. Papanikolopoulos, "Combining multiple tracking modalities for vehicle tracking at traffic intersections," in *Proc. IEEE ICRA*, 2004, vol. 3, pp. 2303–2308.

[162] M. Molinier, T. Häme, and H. Ahola, "3D-connected components analysis for traffic monitoring in image sequences acquired from a helicopter," in *Image Analysis*. Berlin, Germany: Springer-Verlag, 2005, pp. 141–150.

[163] A. Barth and U. Franke, "Tracking oncoming and turning vehicles at intersections," in *Proc. 13th IEEE ITSC*, Sep. 2010, pp. 861–868.

[164] I. E. Paromtchik, M. Perrollaz, and C. Laugier, "Fusion of telemetric and visual data from road scenes with a lexus experimental platform," in *Proc. IEEE IV Symp.*, 2011, pp. 746–751.

[165] M. Muffert, T. Milbich, D. Pfeiffer, and U. Franke, "May I enter the roundabout? A time-to-contact computation based on stereo-vision," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 565–570.

[166] H. Y. Cheng and J. N. Hwang, "Multiple-target tracking for crossroad traffic utilizing modified probabilistic data association," in *Proc. IEEE ICASSP*, 2007, vol. 1, pp. I-921–I-924.

[167] T. Sayed, G. Brown, and F. Navin, "Simulation of traffic conflicts at unsignalized intersections with TSC-sim," *Accid. Anal. Prev.*, vol. 26, no. 5, pp. 593–607, Oct. 1994.

[168] S. R. Perkins and J. L. Harris, *Traffic Conflict Characteristics-Accident Potential at Intersections*. Washington, DC, USA: Highway Research Record, 1968.

[169] N. Saunier, T. Sayed, and C. Lim, "Probabilistic collision prediction for vision-based automated road safety analysis," in *Proc. IEEE ITSC*, 2007, pp. 872–878.

[170] S. Lefèvre, C. Laugier, and J. Ibañez-Guzmán, "Evaluating risk at road intersections by detecting conflicting intentions," in *Proc. IEEE/RSJ Int. Conf. IROS*, 2012, pp. 4841–4846.

Sokèmi René Emmanuel Datondji received the Bachelor's degree in information technology from Aarhus University, Aarhus, Denmark, in 2013 and the M.Eng. degree in embedded systems from École Supérieure d'Ingénieurs en Génie Électrique, Rouen, France, in 2014. He is currently working toward the Ph.D. degree in computer science from University of Rouen, Rouen. He is affiliated with both CEREMA and LITIS Laboratory. His research interests include computer vision, intelligent transportation systems, and pervasive computing.

Yohan Dupuis received the M.Sc. degree in electrical engineering from Union Graduate College, Schenectady, NY, USA; the M.Eng. degree from École Supérieure d'Ingénieurs en Génie Électrique, Rouen, France, in 2009; and the Ph.D. degree in computer science from Université de Rouen, Rouen, in 2012. He is currently a Research Engineer with CEREMA. His research interests focus on perception for vehicle–infrastructure interaction understanding.

Peggy Subirats received the M.Sc. degree in control engineering and applied informatics from École Centrale de Nantes, Nantes, France, and the M.Eng. degree and the Ph.D. degree in control engineering and applied informatics from University of Nantes, Nantes, in 2003 and 2006, respectively. She is currently an Engineer with CEREMA. Her research interest includes driver/vehicle/infrastructure interaction analysis.

Pascal Vasseur received the M.S. degree in system control from Université de Technologie de Compiègne, Compiègne, France, in 1995 and the Ph.D. degree in automatic control from Université de Picardie Jules Verne, Amiens, France, in 1998. Between 1999 and 2010, he was an Associate Professor with Université de Picardie Jules Verne. He is currently a Full Professor with Université de Rouen, Rouen, France, and is a member of the LITIS Laboratory. His research interests include computer vision and its applications to mobile and aerial robots.