IET The Institution of Engineering and Technology WILEY

REVIEW

Vision-based vehicle speed estimation: A survey

David Fernández Llorca[1,2] 🆔 | **Antonio Hernández Martínez**[1] | **Iván García Daza**[1]

[1] Computer Engineering Department, University of Alcalá, University Campus, Alcalá de Henares, Madrid, Spain

[2] European Commission, Joint Research Center, Seville, Spain

Correspondence
David Fernández Llorca, Computer Engineering Department, University of Alcalá, University Campus, Alcalá de Henares, 28805, Madrid, Spain.
Email: david.fernandez-llorca@ec.europa.eu

Funding information
Community Region of Castilla la Mancha, Spain, Grant/Award Number: CLM18-PIC-051; Community Region of Madrid (Spain), Grant/Award Number: S2018/EMT-4362 SEGVAUTO 4.0-CM; Ministerio de Ciencia e Innovación, Grant/Award Number: DPI2017-90035-R

Abstract
The need to accurately estimate the speed of road vehicles is becoming increasingly important for at least two main reasons. First, the number of speed cameras installed worldwide has been growing in recent years, as the introduction and enforcement of appropriate speed limits are considered one of the most effective means to increase the road safety. Second, traffic monitoring and forecasting in road networks plays a fundamental role to enhance traffic, emissions and energy consumption in smart cities, being the speed of the vehicles one of the most relevant parameters of the traffic state. Among the technologies available for the accurate detection of vehicle speed, the use of vision-based systems brings great challenges to be solved, but also great potential advantages, such as the drastic reduction of costs due to the absence of expensive range sensors, and the possibility of identifying vehicles accurately. This paper provides a review of vision-based vehicle speed estimation. The terminology and the application domains are described and a complete taxonomy of a large selection of works that categorizes all stages involved is proposed. An overview of performance evaluation metrics and available datasets is provided. Finally, current limitations and future directions are discussed.

1 | INTRODUCTION

The ability to provide an accurate estimate of the speed of road vehicles is a key feature for Intelligent Transport Systems (ITS). This requires tackling problems such as synchronized data recording, representation, detection and tracking, distance and speed estimation etc., and it can be addressed using data from sensors of different nature (e.g. radar, laser or cameras) and under different lighting and weather conditions.

This is a well-known topic in the field with considerable contributions from both academia and industry. The most important application domains are speed limit enforcement, traffic monitoring and control, and, more recently, autonomous vehicles. However, it is important to note that the nature of the problem of speed detection differs in a not insignificant way, depending on whether it is done from the infrastructure (fixed sensors) or from a mobile platform (intelligent/autonomous vehicles or mobile speed cameras). Our study focuses on fixed systems from the infrastructure, although many of the presented methodologies can be easily applicable to the case of mobile sensors by compensating the speed of the ego-vehicle.

The requirements concerning the uncertainty and robustness on the vehicle speed estimation from the infrastructure depends on the application scenario. On the one hand, for speed enforcement applications, extreme maximum permissible error standards are required by national or international certification agencies [1], which is more than reasonable considering that these automatic systems are used to fine, or even imprison, drivers who exceed the maximum speed limit. This implies that the most common sensors used for speed measurement in these applications are based on high-precision high-cost range sensors, such as radar (Doppler effect) and laser (time of flight), or, in some cases, accurate on-pavement sensors (e.g. inductive or piezoelectric loop detectors) embedded in pairs under the road surface. The use of visual cues to measure physical variables related with the motion of the vehicles has been rarely proposed. Video cameras usually play a secondary role as a means of capturing a human-understandable representation of the scene, allowing the visual identification of the vehicles, and serving as evidence for speed limit enforcement. On the other hand, for traffic monitoring, forecasting and control applications, there is no regulatory framework or scientific consensus

on the accuracy and uncertainty of speed estimation [2]. The requirements regarding sensors are therefore reduced, and it is more common to see the use of less accurate sensors [3] including traffic cameras [4]. In any case, it is obvious that the higher the accuracy of the speed estimation, the better the forecasting and control approaches will be.

Recent advances in computer vision techniques have led to a significant increase in the number of works proposing the use of vision as the only mechanism for measuring vehicle speed. The challenge of making accurate estimations of distances and speeds arises from the discrete nature of video sensors that project the 3D world into a 2D discrete plane. This intrinsic limitation results in a digital representation whose accuracy follows an inverse-square law, that is, its quantity is inversely proportional to the square of the distance from the camera to the vehicle. Despite these limitations, the potential benefits of using video cameras are remarkable, since they are a cost-efficient solution compared with range sensors such as microwave Doppler radars or laser-based devices. The possibility of using already installed traffic cameras without the need to integrate new sensors is another advantage. Also, the visual cues from these sensors can be used to address other problems such as automatic identification of the vehicle (license plate, make, model and color), re-identification in different places, and if the resolution allows, the identification of occupants, or seat belt use.

The scope of this survey is vision-based vehicle speed detection from the infrastructure, including fixed point speed cameras and traffic cameras, and excluding mobile speed cameras or onboard cameras for intelligent vehicle applications. Despite the potential impact, and the high number of contributions available, there is still no detailed survey of this topic. We describe the terminology and the application domains. We survey a large selection of works and propose a complete and novel taxonomy that categorizes all the components involved in the speed detection process. Performance evaluation metrics and available datasets are examined. Finally, current limitations are discussed and major open challenges for future research are outlined. We hope that this work will help researchers, practitioners and policy makers to have a detailed overview of the subject that will be useful for future work.

The structure of this paper is as follows: we present the terminology and describe the application domains in Section 2, we present a general description of the taxonomy in Section 3, and we review and analyze each of its components in Sections 4–7. The description of the available datasets and evaluation metrics are provided in Section 7. Discussion of the current state of the art and future trends are provided in Section 8, and, finally, Section 9 concludes the presented work.

2 | OVERVIEW AND TERMINOLOGY

The main components of a fixed vision-based vehicle speed measurement system are summarized in Figure 1. Four levels of abstraction are considered, which are linked to the main terminology in what follows:
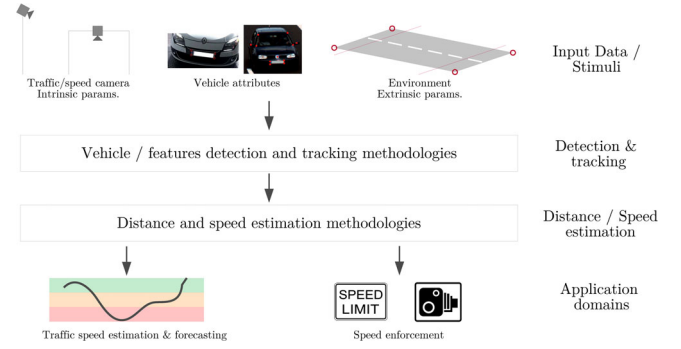

FIGURE 1 Main high-level components of a vision-based vehicle speed estimation system: input data, object detection and tracking, distance and speed estimation, and outcome

2.1 | Input data / stimuli

Vision-based speed detection approaches are based on input data from video cameras. For each vehicle there will be a sequence of images from the first appearance to the last one. The number of images available will depend on the camera pose with respect to the road, the focal length, the frame rate, and the speed of the vehicle. Two different types of video cameras can be considered:

- *Traffic cameras*: also referred as traffic surveillance cameras or traffic CCTV. They are used to monitor traffic flow, congestion and accidents both manually (visual inspection by a human operator) or automatically. In most cases, these are fixed cameras located far from the traffic flow, infrastructure- or drone-based.
- *Speed cameras*: also referred as traffic enforcement camera. They are generally considered to be a device for monitoring the speed of vehicles independently of the sensor used to measure the speed (radar, laser or vision). Even for radar- or laser-based systems, there is always a camera to record an image of the vehicle, and hence the term *camera*. In this case, we use the term speed camera in the most literal way, that is, vision-based systems for measuring speed. Their location is usually closer to the traffic flow than the traffic cameras.

Other forms of input data include vehicle attributes such as vehicle type, keypoints, license plate size etc. Camera calibration plays a key role in providing both intrinsic and extrinsic parameters. Prior knowledge of the size of the road section reveals information that is critical to calculating the extrinsic relationship between the road and the camera or even the speed of the vehicles.

2.2 | Detection and tracking

The vehicles, or some of their representative features, must be detected in all the available images. Tracking the vehicle or its features over time is essential to obtain speed measurements. Different approaches can be applied to address both tasks.

FIGURE 2 Examples of images from different application domains. Left column: traffic and drone cameras. Middle column: speed cameras. Right column: autonomous vehicles context. Images obtained from [88, 100, 105] and [130]

2.3 | Distance and speed estimation

Speed estimation intrinsically involves the estimation of distances with associated timestamps. Different methods exist to calculate the relative distance of vehicles from some global reference, as well as different approaches to finally calculate the speed of the vehicle.

2.4 | Application domains

Vision-based vehicle speed estimation is a fundamental task for traffic monitoring, forecasting and control applications, speed enforcement, and, in addition, autonomous or intelligent vehicles (different examples are depicted in Figure 2).

2.4.1 | Traffic monitoring and control

The knowledge of the speed of all the vehicles in multiple road segments allows to know, and even forecast, the traffic state of the road (flow, density and speed), which can be used for a very wide set of different applications such as adaptive traffic signal control, real-time traffic aware navigation, traffic jams, congestion and accident detection, as well as other off-line operation and planning scenarios such as infrastructure enhancement.

2.4.2 | Speed enforcement

Speed has been identified as a key risk factor in road traffic injuries, increasing both the risk of road traffic crashes and the severity of the consequences [5]. And speed enforcement is a key road safety measure that leads to a reduction in accidents [6]. As suggested in [7], in the vicinity of speed cameras the reduction of speeding vehicles and crashes can reach up to 35% and 25% respectively. Moreover, the higher the intensity or level of enforcement, the larger the accident reduction [8].

Depending on the country, speed offenders can be fined, lose their driver's license, or even go to jail. Therefore, speed estimation technology must be extremely accurate. National and international metrology or measurement authorities impose high standards of accuracy that must be met to allow the use of speed cameras to enforce speed limits. The maximum permissible errors (MPE) are usually defined as:

$$MPE \leq \begin{cases} \pm N, & \text{if } S \leq 100 \text{ km/h} \\ < N\%, & \text{if } S > 100 \text{ km/h} \end{cases} \quad (1)$$

where S is the actual speed of the vehicle, and N is a value usually defined between 1 and 3 depending on the country. Designing vision-based vehicle speed estimation systems that meet these accuracy requirements is certainly a complex challenge due to the intrinsic limitations of vision sensors for measuring distances.

2.4.3 | Autonomous vehicles

The ability to estimate the relative distances and speeds of surrounding vehicles is essential for advanced driver assistance systems and automated driving. From simple Adaptive Cruise Control (ACC) systems [9], through Traffic Jam Assist and Chauffeur systems that combine ACC Stop & Go [10] and Lane Keeping Assist, to sophisticated Highway Chauffeur or Autopilot system which include complex maneuvers such as overtaking [11], the accurate localization of target vehicles and the estimation of their relative speed are fundamental variables used by low-level control systems. Vision-based approaches have been widely addressed to deal with vehicle detection, using flat-world assumption, stereo-vision or by combining cameras with range-based sensors such as radar or LiDAR. Although vehicle speed estimation for autonomous vehicles is out of the scope of this survey, most of the analyzed approaches can be extended to be used on-board of mobile platforms.
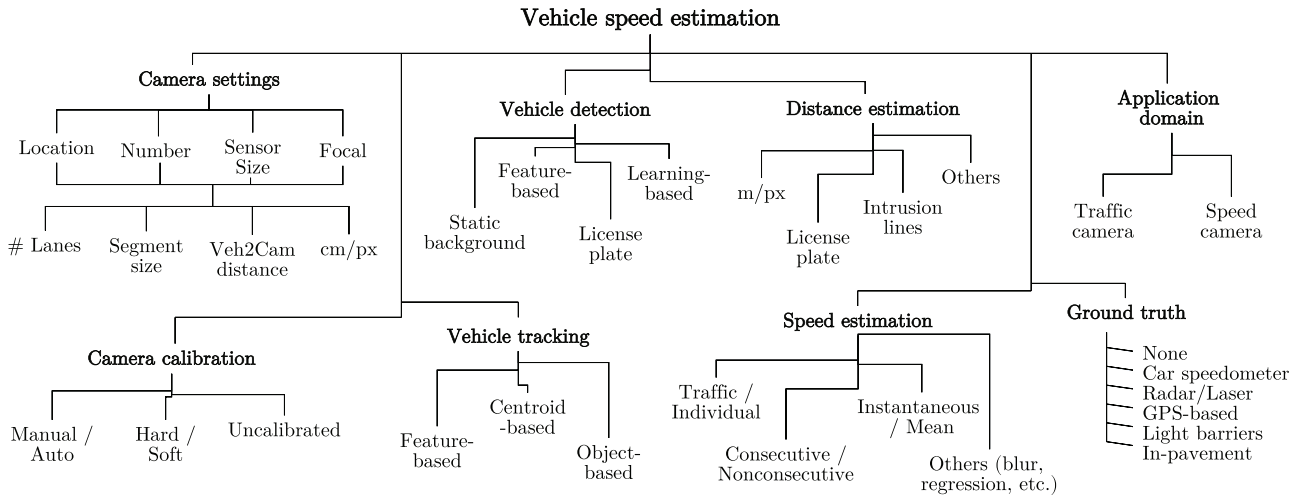
FIGURE 3 Proposed vision-based vehicle speed estimation taxonomy

3 | TAXONOMY

In this survey, more than 135 papers related with vision-based vehicle speed detection have been evaluated. To provide a consistent numbering of bibliographic references, these are provided in temporary order by year. The proposed taxonomy breaks down the problem of speed detection, including all the components, approaches and experimental details involved (see Figure 3 for a general overview). We start with the input stimuli and explain how camera settings and calibration procedures affect the subsequent steps. Next, we analyse all the steps involved, from vehicle detection and tracking to distance and speed estimation. Finally, we discuss the different approaches to evaluate the accuracy of the speed measurements.

On the one hand, the taxonomy presented is intended to serve as a basis for understanding what the main elements and issues are from a high-level perspective (e.g. practitioners and policy makers). On the other hand, it is intended to serve as a tool for the development of future approaches for future research.

3.1 | Application domain

The first categorization is directly related to the attempt to devise an accurate system to be used for speed enforcement (**speed camera** application domain) [13, 17, 18, 32, 33, 39, 41–43, 51, 53, 54, 61, 65–68, 70, 72, 74, 76–78, 80, 82, 84–86, 94, 100–102, 107, 108, 111, 112, 114, 119, 122–125] or a less accurate approach that can still be used for traffic speed estimation (**traffic camera** application scenario) [12, 14–16, 19–31, 34–38, 40, 44–50, 52, 55–60, 62–64, 69, 71, 73, 75, 79, 81, 83, 87–93, 95–99, 103–106, 109–111, 113, 115–118, 120, 121, 126–129]. Although in some cases it is claimed that accurate speed detection is targeted, some aspects such as the low resolution of the camera, a limited experimental evaluation or the absence of an appropriate ground truth make us to categorize these manuscripts as traffic speed systems.
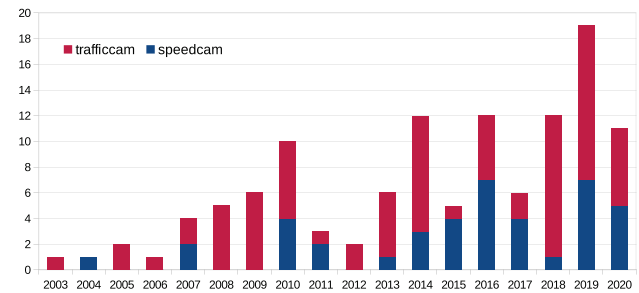

FIGURE 4 Publication trends in the state of the art by application domain.

With this categorization, in Figure 4 we present the distribution of the papers published by year. As can be seen, the general trend in the number of publications is upwards, with the number of proposals focusing on accurate speed detection for speed enforcement being much lower than the number of proposals focusing on traffic speed detection.

3.2 | Camera settings

This category involves the camera intrinsic (sensor size and resolution, focal length) and extrinsic (location with respect to the road plane, including drone-based cameras) parameters, and the number of cameras (mono, stereo or multiple). Depending on these parameters, the real-world scenario projected on the image plane can represent one or several lanes along a short or large road segment. Depending on this configuration, and the relative position of the vehicles with respect to the camera, we finally have one of the most important variables: the **meter-to-pixel ratio**, that is, the portion of the road segment represented by each pixel. The lower this ratio is, the greater the accuracy in estimating distance and speed. Due to the perspective projection model, this ratio is proportional to the square of the distance from the camera, which means that those measurements

taken at long distances have very poor accuracy. Note that we omit integration details such as the exact location of the camera, anchoring systems, type of gantry, pole or bridge etc. These features are rarely provided in the reviewed works.

3.3 | Camera calibration

An accurate estimation of the camera intrinsic and extrinsic parameters is required to provide measurements in real-world coordinates. The most common approach is to perform a **soft calibration**, that is, intrinsic parameters are calibrated in the lab or approximated using the sensor and lens characteristics, and extrinsic parameters (rigid transformation between the road plane and the camera) are obtained using a manual [17, 28, 31, 35, 37, 47, 54, 59, 66, 67, 70, 74, 76, 80, 83, 86, 91, 92, 94, 100, 109, 110, 117, 119, 121, 125, 128, 129] or an automatic procedure [77, 90]. **Hard calibration** involves to jointly estimate both intrinsic and extrinsic parameters with the camera already installed. It can also be performed either manually [20, 38, 58] or automatically [12, 14, 18, 19, 23, 33, 78, 85, 88, 101, 108]. In some limited cases, camera calibration is somehow neglected [13, 15, 16, 21, 22, 24–26, 30, 32, 34, 39–41, 43, 44, 46, 48–53, 55, 56, 60–65, 69, 71, 73, 81, 84, 95, 103–107, 111, 113, 118, 120, 124].

3.4 | Vehicle detection

Since the cameras are mostly static (except for the drone cameras), in most cases, vehicle detection is addressed by modeling and subtracting the background [12–15, 18, 20–28, 30, 32, 34–37, 40, 42–44, 47, 48, 50, 51, 53–55, 61–63, 65, 71–75, 81, 83, 84, 86, 92, 93, 95, 102, 103, 108, 110–113, 117, 118, 121]. Other approaches are feature-based, for example, detecting the license plate [31, 41, 49, 60, 66, 68, 70, 76, 77, 80, 82, 94, 122, 125] or other features of the vehicle [38, 39, 45, 49, 57–59, 69, 71, 79, 87, 91, 95, 96, 98, 99, 110–112, 116, 117, 123]. The use of learning-based methods to recognize vehicles in images has been growing recently [19, 85, 88–91, 98, 100, 101, 104–106, 115, 120, 126, 127, 129].

3.5 | Vehicle tracking

The ability to have a smooth and stable trajectory of the vehicles is a key issue to deal with vehicle speed detection. We can classify vehicle tracking into three different categories. First, feature-based [17, 19, 20, 22, 31, 33, 37, 38, 45, 51, 53, 55, 58, 59, 61, 62, 69, 71, 74, 79, 85, 88, 91–93, 97, 112], which tracks some set of features from the vehicle (e.g. optical flow). Second, those methods focused on tracking the centroid of the blob or the bounding box of the vehicle [25, 34, 35, 40, 48, 50, 52, 56, 65, 72, 73, 75, 83, 86, 95, 103, 106, 111, 118, 121, 122]. Finally, those approaches that focus on tracking the entire vehicle [15, 16, 18, 26–30, 32, 36, 42–44, 47, 54, 57, 63, 78, 81, 84, 89, 91, 98, 102, 104, 105, 108, 109, 111, 113, 114, 120, 125, 128, 129], or a spe-

cific part of it (such as the license plate [41, 49, 60, 66, 68, 70, 76, 77, 80, 82, 94, 100, 125]).

3.6 | Distance estimation

For monocular systems, the estimation of the distance of the vehicles is usually computed with a set of constraints such as flat road assumption, including methods based on the homography [15, 16, 18, 20, 22, 34, 37, 52, 53, 61, 66, 94, 98, 99, 110] and the use of augmented intrusion lines, patterns or regions [19, 20, 23, 24, 26, 32, 38, 47, 79, 86, 87, 107, 124], or by using prior knowledge about the actual dimensions of some of the objects (e.g. the license plates [41, 49, 60, 66, 68, 80, 125] or the size of the vehicles [42, 62, 78, 103]). These restrictions are alleviated when using stereo vision [42, 60, 61, 74, 80, 92, 100, 112].

3.7 | Speed estimation

In a few cases, the problem of detecting the speed of vehicles is posed as a problem of detecting the speed of traffic on a segment of the road, that is, directly obtaining an average road speed value [12, 17, 20, 23, 32, 34, 38, 40, 42, 51, 53, 54, 58, 61, 69, 77, 79–81, 84, 85, 94, 102, 105, 110, 113, 121]. However, in most cases the detection is carried out at the level of individual vehicles. Prior knowledge about the frame rate of the camera or accurate timestamps per each image are needed to compute the time between measurements. The use of consecutive [14, 18, 19, 22, 25, 33, 35–37, 41, 46, 48, 50, 55, 65, 67, 68, 71–76, 78, 82, 83, 85, 88, 89, 92, 96, 100, 103, 120, 129] or non-consecutive [15, 24, 26, 28, 30, 31, 43, 44, 47, 49, 52, 59, 66, 70, 71, 86, 87, 93, 95, 99, 106–108, 111, 118, 124, 125, 128] images to estimate speed is a fundamental variable that has a considerable impact on accuracy. How to integrate all available measurements (instantaneous, mean, optimal etc.) is also a key factor affecting the final accuracy of the measurement.

3.8 | Ground truth generation

To evaluate the accuracy of the different vehicle speed estimation methodologies, an appropriate ground truth of the actual vehicle speed is required. Surprisingly, a considerable number of works presented their results without comparing them with reference obtained with a higher precision system. This raises questions about the validity of the results presented.

As can be observed in Figure 3, the technologies used to compute the ground truth include the following: car speedometers, range-based technologies such as radar or laser, GPS-based including standard GPS, GPS odometers, and DGPS, light barriers and in-pavement sensors such as piezoelectric or inductive.

Even though it is more of a technological problem than a scientific one, the synchronization of the data from the cameras and the other sensors is a critical and complex issue [80]. The sensors are usually located in different places, with independent recording systems, and with different sampling frequencies.
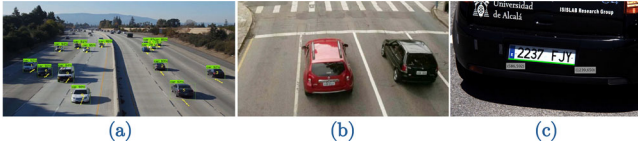
FIGURE 5 Examples of camera settings. (a) Low-medium focal length, very far relative distances. Multiple lanes, both directions. Very high meter-to-pixel ratio [88]. (b) Medium focal length, medium relative distances. Multiple lanes, one direction. Medium meter-to-pixel ratio [82]. (c) High focal length, low relative distances. One lane, one direction. Low meter-to-pixel ratio [80]

4 | CAMERA SETTINGS AND CALIBRATION

4.1 | Camera settings

The camera settings have a direct impact on the accuracy of the speed estimation method. The first categorization considered is the camera position, including drone cameras [25, 30, 57, 62, 64, 69, 96, 104–106, 126], and traffic cameras that can be classified roughly according to their height as distant (≥ 5 m) or close (< 5 m) traffic cameras. In some cases, the cameras are located on the side of the road, providing side views of the vehicles [13, 21, 26, 33, 39, 73, 112, 118]. Most works are based on monocular systems, but we can also find a few approaches based on stereo [42, 60, 74, 100, 112] and non-stereo multi-camera [61, 80] setups. Although most approaches are based on conventional cameras with CCD or CMOS sensors, we can find some works using event-based cameras [16, 17], based on bio-inspired sensors [147].

Considering camera resolution one can expect to have a progressive increase of the pixel resolution over time, mainly due to advances in hardware. However, this is not the case, and a considerable number of modern works (since 2010 we have [32–35, 42–45, 52, 54, 67, 71–73, 79, 83, 87, 97, 103, 117]) with a resolution lower than or equal to 640×480 pixels (VGA). Resolution not only affects the accuracy of image processing detection techniques, but also the accuracy of distance estimation. The higher the resolution, the lower the meter-to-pixel ratio.

The focal length is also a fundamental parameter that is usually linked to the height of the camera, the length of the road segment and the number of lanes covered by the field of view. As explained in Sections 6.1 and 6.3, for a specific camera height, the larger the focal length the higher the accuracy of the distance and speed measurements.

Most works use a low-medium focal length (≤ 25 mm) since they are designed to cover multiple lanes and a large road stretch. Only a few works implicitly or explicitly highlight the need for a long focal length to increase the accuracy of the speed estimation, even if only one lane is covered by the system [23, 31, 53, 61, 68, 70, 76, 77, 80, 82, 107, 124].

Therefore, depending on the camera intrinsics and extrinsics (location w.r.t. the road) parameters we finally have different scenarios projected into the image plane. As depicted in Figure 5, we can roughly define three type of scenarios:

- *High meter-to-pixel ratio* (Figure 5a): obtained with low-medium camera resolution, low focal lengths, high camera height, covering a long road segment, multiple lanes and both directions. Some examples are [12, 15, 19, 22, 27, 36, 48, 74, 75, 81, 88, 121, 128, 129].
- *Medium meter-to-pixel ratio* (Figure 5b): obtained with medium-high camera resolution, medium focal lengths, medium camera height, covering a medium road segment, multiple lanes and only one direction. Some examples are [18, 41, 42, 45, 54, 61, 76, 82, 93, 108].
- *Low meter-to-pixel ratio* (Figure 5c): obtained with high camera resolution, high focal lengths, low camera height, covering a short road segment, one or two lanes and only one direction. Some examples are [23, 31, 53, 68, 70, 77, 80, 107, 124].

4.2 | System calibration

Accurate camera calibration is of utmost importance for vision-based distance and speed estimation of objects. Whether assuming that vehicles are moving along a flat road for monocular systems or using a stereo configuration, the ability to obtain accurate real-world measurements from pixel coordinates considerably depends on accurate system calibration. For monocular systems, the pinhole camera model provides a general expression to describe the geometric relationship between a 2D pixel in the image plane (u, v) and a 3D real-world coordinate system (X_w, Y_w, Z_w), in this case, placed on the road plane and assuming a flat road shape:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \qquad (2)$$

where f_x and f_y are the focal distances along the x and y axes of the camera, respectively, c_x and c_y represent the image optical center. These are the intrinsic parameters that can be represented in the intrinsic matrix K, which assumes that the distortion of the image has been appropriately corrected (note that the use of large focal lengths considerably reduces the effect of the distortion). The rotation and translation parameters can be grouped in a 3×3 rotation matrix R and a 3×1 translation vector T.

The monocular camera calibration problem is the problem of estimating K, R and T. As stated in Section 3.3, and depicted in Figure 6, two main approaches are applied when dealing with camera calibration for vehicle speed estimation. First, **soft** calibration (Figure 6a), which estimates K in the lab using calibration patterns (e.g. chessboard) or assumes the parameters provided in the datasheets (sensor resolution, pixel size and focal length), and then computes $[R, T]$ on site. Second, **hard** calibration (Figure 6.b) which calculates K and $[R, T]$ on site.

One of the most common ways to compute the camera's extrinsic parameters in the operation environment is using vanishing points [12, 14, 15, 19, 22, 23, 27, 33, 42, 85, 88, 89, 91, 108]. The set of parallel lines in the 3D real-world coordinate
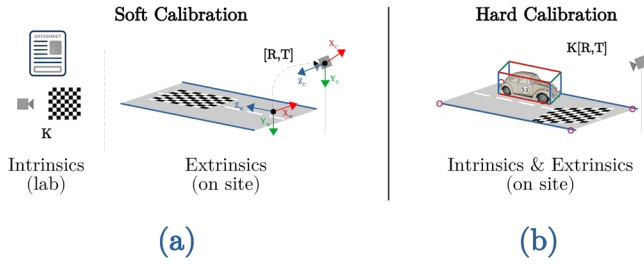
FIGURE 6 Camera calibration procedures. (a) **Soft calibration**. Intrinsic parameters (focal distance, optical center and distortion coefficients) are computed in the lab. Then, for a fixed focal length, the extrinsic parameters are computed on-site. (b) **Hard calibration**. Both intrinsic and extrinsic parameters are computed on site. In both cases the procedures can be manual or automatic

system intersect at a unique 2D point when projected to the image plane. This point is commonly referred to as the vanishing point. The obvious set of lines in this case are the road markings (including solid and dashed lane lines) and the road edges that are usually of high contrast. The location of the vanishing point allows the computation of the rotation matrix R. The translation matrix T is then obtained using knowledge of real-world dimensions of some object or region in the image. Another common approach to perform calibration of extrinsic parameters is using known geometric static features present on the road plane and then compute the homography transformation which includes rotation, translation and scale [31, 38, 47, 66, 67, 70, 72, 75, 79, 82, 86, 87, 98, 99, 110, 113, 119, 121, 128, 129]. The dimensions, distances, lengths etc., of these static features have to be known in advance. This may involve the need to stop traffic to access the operating environment and take direct measurements or use some indirect method (e.g. using laser scanners, or even Google Maps).

Hard calibration to get both camera's intrinsic and extrinsic parameters can be addressed by using standard (e.g. chessboard) or ad-hoc calibration patterns [17, 18, 37, 58, 59, 68, 76, 80, 100] which are placed on the road plane. The main drawback of this technique is that it requires stopping traffic on the road or at the lane level temporarily. Only a few approaches tackle the calibration of both intrinsic and extrinsic parameters in an automatic fashion. The most common approach is to calculate two or three orthogonal vanishing points from two or three sets of parallel lines perpendicular to each other [133]. As described in [134], these vanishing points can be extracted from both static and moving (vehicles in this case) elements present on the scene [23, 33, 85]. In [77] the license plate tracked over time is used as a calibration pattern to hard calibrate the system. Most sophisticated approaches combine fine-grained classification to determine the specific vehicle make and model [135] (and therefore, their real dimensions) with vehicle keypoint/pose detection [136] to perform either soft [90] or hard [101] system calibration.

Regardless of the method used in the calibration, either to calculate the translation vector, or the complete homography, it is necessary to know a priori the dimensions of some fea-

ture/object in the real world. In fact, computing the so-called **scale factor (m/px)** to transform from pixel to real-world coordinates is one of the key problems to be solved when dealing with distance and speed estimation using monocular systems. The most common features are the road/lane width or the length of a section/region manually measured [19, 24, 38, 47, 54, 64, 66, 67, 72–75, 79, 82, 84, 87–89, 98, 101, 107, 113, 128, 129], the size of a previously known object such as vehicles [27, 62, 78, 85, 90, 103, 105, 106, 110, 117] or license plates [60, 66, 68, 70, 80, 94, 133], and the length and frequency of the lane markings [14, 20, 22, 23, 26, 61, 65, 78, 108, 119, 121].

Finally, we have to consider stereo-based approaches, which are hard calibrated in the lab or on-site using standard stereo calibration techniques and patterns [60, 74, 92, 100, 112]. In addition to the intrinsic matrix K per each camera, stereo systems requires the computation of the fundamental matrix which includes the extrinsic transformation from one camera to the other one. Thanks to the ability of stereo systems to obtain 3D real-world measurements from 2D pixels in both images (after solving the correspondence problem) there is no need to compute the extrinsic transformation between the road plane and the cameras, but, in any case, it can be computed manually [137] or automatically computed [138].

5 | VEHICLE DETECTION AND TRACKING

This section provides a brief overview of the different vision-based vehicle detection and tracking approaches used to measure vehicle speed. However, the techniques applied to address vehicle detection and tracking as the first stages of vehicle speed detection systems are practically the same as those used for traffic surveillance applications. Therefore, we refer to [4] for a more in-depth and specific analysis of the different approaches.

5.1 | Vehicle detection

5.1.1 | Static background

With the exception of drone-based systems, most approaches are based on static cameras, making the vehicle detection task much easier thanks to the presence of a static background. We can even find a considerable number of works applying the simplest frame-by-frame approach followed by a thresholding method to perform image segmentation [14, 32, 34, 36, 37, 40, 42, 46, 54, 73, 103]. The most common approach to perform vehicle detection is based on background subtraction [12, 13, 15, 18, 20–28, 30, 36, 37, 43, 44, 47, 48, 50, 51, 53, 55, 61–63, 65, 71, 72, 74, 75, 81, 83, 84, 86, 92, 93, 95, 102, 108, 110, 112, 113, 117, 118, 121], followed by some morphological operations and a blob analysis method. Different methodologies are used to perform background subtraction, including gray- and

color-based approaches, Gaussian Mixture Models [23, 61, 84, 102], adaptive background modeling [51, 111] etc. In some cases, frame-by-frame or background subtraction is combined with some feature detector (e.g. edges, corners, or KLT features) [40, 48, 50, 71, 110].

5.1.2 | Feature-based

Some approaches are based on the detection of different type of features grouped in regions within the vehicle area, for example, after background subtraction [38, 39, 57, 99, 110, 111, 116], as well as simple methods that rely on edges [49, 87, 95], gray-level features [45, 117, 123], binary features or patterns [71, 87, 96], corners [79], SIFT/SURF features [62, 112] or KLT features [58, 59, 69, 91].

5.1.3 | License plate detection

Although the appearance of the license plates may vary a little depending on the country and the type of vehicle, in general it is a homogeneous and distinctive element of any vehicle. Indeed, it is practically impossible for a detected plate not to correspond to a vehicle, and it is very unlikely that a vehicle will transit without a plate, as this is something that is criminally punishable throughout the world. This makes the license plates a very attractive target for vehicle detection systems [31, 41, 49, 60, 66, 68, 70, 76, 77, 80, 82, 94, 122, 125]. Therefore, License Plate Recognition (LPR) systems can be applied when camera resolution is large enough, that is, from medium to low meter-to-pixel ratios. LPR systems are usually divided into at least three steps: detection or localization, character segmentation and character recognition. In this case, to detect a vehicle, it is only necessary to apply the first step, and there are many methods for this (we refer to some of the multiple specific surveys on this topic [139, 140]).

5.1.4 | Learning-based

As is the case in other fields of application, learning-based approaches are recently gaining ground thanks to the availability of data and the advance in machine learning. Although the first approach based on learning was found in 2007 [19], using Boosting type methodologies (Adaboost, also used in [120], which is the only one not based on deep learning), it is not until 2017 that we see consolidated use of this type of methodologies in all the application domains, including Faster R-CNN [85, 88, 90, 98, 101, 106, 128] an the extended version Mask R-CNN [89], SSD (Single Shot MultiBox Detector) [100, 128] and different versions of the YOLO (You Only Look Once) detector, that is, YOLO [115], YOLOv2 [128, 129] and YOLOv3 [104, 105, 126]. In [91] transfer learning is proposed to fine-tune a vehicle detection model for their specific traffic scenario.
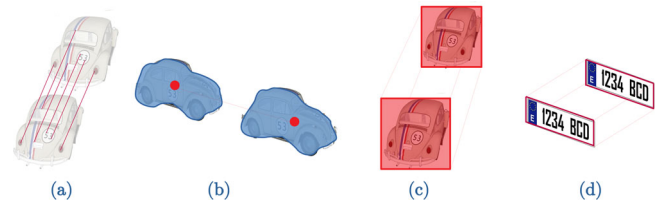

FIGURE 7 Vehicle tracking approaches. (a) Feature-based. (b) Blob centroid tracking. (c) Bounding box tracking. (d) License-plate-based

5.2 | Vehicle tracking

Speed detection is a multi-frame problem. Once the vehicles or some parts/features of them are detected, a tracking method is usually applied to filter and smooth noisy detections. Tracking implicitly involves solving the data association problem (e.g. Hungarian algorithm), especially for cases with multiple objects [148]. In general terms, two main dimensions can be considered when describing the different tracking approaches. First, in terms of the type of element and/or parts of it being tracked (approaches), and second, in terms of the specific methodology used for carrying out the tracking (methodologies).

5.2.1 | Approaches

A visual representation of the main vehicle tracking approaches is presented in Figure 7. As can be observed we can classify the different approaches into four main categories. First, **feature-based** approaches which pose the vehicle tracking problem as a feature tracking problem. In most cases, the features are either binary or gray level values within the vehicle region after applying background subtraction or a learning-based method [20, 37, 38, 45, 51, 61, 74, 85, 92, 93]. KLT features or corners are usually considered when tracking is posed as an optical flow problem, analyzing motion vectors within the vehicle region [19, 33, 53, 55, 58, 59, 69, 79, 88, 91, 97]. Other tracked features are binary patterns [71], edges [55, 79] or SIFT/SURF descriptors [62, 75, 112].

Second, we can find methods focusing on **tracking the centroid** of the region representing the vehicle, which can be considered as the contour (either convex hull or concave) [48, 65, 72, 73, 111, 121], or some model such as the convex hull of the contour [40], an ellipse [75] or the bounding box fitted to the detected blob [34, 35, 50, 56, 83, 86, 95, 103, 106, 118]. In any case, the use of the centroid as a representative state of the vehicle is very unreliable. On the one hand, most of the blob detection approaches are not very accurate and the vehicle contour may vary in the sequence due to multiple factors (shadows, overlaps, close objects etc.). This makes the centroid rarely representing the same point of the vehicle in time. But even in those cases in which the blob detection is very precise (e.g. learning-based) the changes of perspective imply considerable variations that affect the location of the centroid. Only in very particular cases where detection is accurate and perspective changes are

not relevant (for example, in drone based systems), is the use of this approach justified.

A third group of approaches are those that use the **entire region of the vehicle** (contour- or bounding box-based) to perform the tracking. This is by far the most widely used approach [15, 16, 18, 26–30, 32, 36, 42–44, 47, 54, 56, 57, 63, 78, 81, 84, 91, 98, 102, 104, 105, 108, 109, 111, 113, 114, 120, 125, 128, 129]. Finally, a number of approaches focus on the **tracking of the license plate** [41, 49, 60, 66, 68, 70, 76, 77, 80, 82, 94, 100, 125]. Somehow this type of approaches do not calculate the speed of the vehicles but of the license plates, something completely equivalent and perfectly valid since the movement of the plate is consistent with that of the vehicle.

5.2.2 | Methodologies

Most of the works address the vehicle tracking problem following a **tracking by detection** methodology [15, 17, 18, 20, 25, 31, 32, 34, 38, 40, 42–45, 47, 48, 50, 51, 55, 56, 62, 71–75, 81, 83, 86, 91–93, 98, 102, 105, 106, 111, 112, 118, 120, 121, 125, 128] by means of template, regions or feature matching/correlation methods, by detecting the intersection of the vehicles over predefined regions or intrusion lines, among others. These methods solve the data association problem (associating a unique identifier to each object for each track), but do not filter the state of the tracked object. A particular case of tracking by detection methods is that of license plates, since they allow the use of Optical Character Recognition (OCR) systems as a support for solving the data association problem, making the tracking system more robust [41, 49, 60, 66, 68, 70, 76, 77, 80, 82, 94, 100, 125].

The use of KLT features for vehicle detection is usually followed by a KLT tracked to detect the optical flow [19, 53, 54, 58, 59, 69, 79, 84, 88, 91, 97, 105]. This method can only be applied when the vehicle displacement between the images is small. As we will see in the Section 6.3, this is not appropriate to obtain accurate speed measurements.

The use of Bayesian filtering is limited to the classic linear Kalman filter [22, 27, 61, 65, 78, 85, 104, 109] which is usually applied considering constant velocity models (less applicable for very large road sections). In [89] Simple Online and Real Time (SORT) (which is based on the Kalman filter) tracking and the extended version DeepSORT are also applied.

6 | VEHICLE DISTANCE AND SPEED ESTIMATION

6.1 | Fundamentals of camera-based distance error

Cameras are sensors that generate a discretized representation of the 3D world projected into a 2D plane. Considering the solid angle between the optical center and a pixel, the area in real-world coordinates, A_w, increases quadratically with distance Z:

$$A_w = \frac{A_{px}Z^2}{f^2},\qquad(3)$$

where f is the focal distance and A_{px} is the area of a pixel. Even with perfect calibration of the intrinsic parameters of the camera, and with a zero pixel localization error, the uncertainty or error of distance measurements also increases quadratically with object distance for both monocular (e.g. with prior knowledge about object width [80] or flat road assumption [131]) and stereo systems [132]. Distance error for monocular systems can be formulated as:

$$Z_{err_m} = \frac{Z^2}{f_x \Delta X} n_m,\qquad(4)$$

where f_x is the focal distance in x axis in pixels, ΔX is the known object dimension (e.g. the width of the car [131] or the license plate [80]), and n_m represents the pixel detection error. A similar expression can be formulated for stereo systems:

$$Z_{err_s} = \frac{Z^2}{f_x B} n_s,\qquad(5)$$

where B is the baseline or distance between the cameras, and n_s is the disparity error.

According to the previous expressions, we can conclude that to minimize the error in the estimation of distances of objects or features of these, it is required to:

- Use telescopic lenses (high focal length), since the error decreases with the focal distance.
- Use large stereo baselines in stereo systems, or use objects of known dimensions as large as possible in monocular systems.
- Avoid taking measurements of vehicles when they are far away from the camera. The closer the object, the smaller the error in distance estimation.

The examples depicted in Figure 5 assume these requirements from less (Figure 5.a) to more (Figure 5.c) severity.

6.2 | Relative distance estimation

6.2.1 | Stereo-based approaches

Measuring distances in real-world coordinates is the most critical problem to be solved for accurate vehicle speed estimation. This task is straightforward when using stereo vision [42, 60, 61, 74, 80, 92, 100, 112]. For each detected vehicle, relative distances can be directly obtained using the disparity values of the pixels contained in the region of the vehicle. However, as established in Section 6.1, the accuracy of the measurements given by the stereo system is compromised by several factors such as the distance of the vehicle with respect to the stereo pair of cameras,
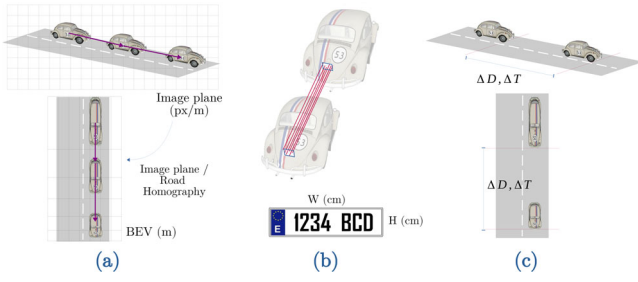
FIGURE 8 Distance estimation approaches: (a) image plane-road plane homography (px/m transformation), (b) prior knowledge of the real-world size of the license plate, the vehicle or any other object; (c) intrusion or augmented lines, vehicle intersection with the road
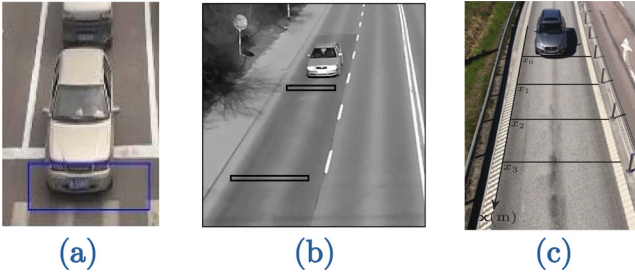

FIGURE 9 Distance estimation based on previous knowledge of the dimensions of a (a) road section [47], (b) two [87] and (c) multiple intrusion or augmented lines [107]

FIGURE 10 Distance estimation based on homography. (a) Calibration using previous knowledge of the real-world dimensions of a road section [53, 66]. (b) Examples of the BEV or image warping [22, 98].

FIGURE 11 Illustration of distance/speed errors in homography-based approaches. (a) Perspective issues in which the distance measured by the camera system $\Delta Z'$ does not correspond to the same point. ΔZ denotes the real distance which cannot be obtained since the contact point is not visible in $t + 1$. This problem also appears when using intrusion lines. (b) Wrong assumption of motion on the road plane. ΔZ denotes the real distance traveled by, for example, the license plate (which takes place at height h), and $\Delta Z'$ corresponds to the wrong estimation of the distance traveled by the license plate after being projected to the road plane using the homography transformation

the baseline between the cameras and the disparity error. The number of proposals using stereo cameras to deal with vehicle distance estimation is somehow limited due to the strong calibration requirements and the fact that most of the traffic cameras already installed are monocular systems.

6.2.2 | Monocular-based approaches

As described in Section 4.2, the previous knowledge of the dimension in real-world coordinates of some features, objects, or parts/sections of the road is a fundamental issue for the estimation of distances using monocular systems. This is usually referred to as the scale factor for transforming pixels to real-world coordinates. Another common requirement is to consider the flat road assumption.

As depicted in Figure 8 we categorize monocular-based distance estimation approaches into three groups. First, based on intrusion or augmented lines or regions [19, 20, 23, 24, 26, 28, 32, 38, 47, 59, 79, 86, 87, 99, 107, 124]. These approaches do not require calibration of the camera system but measure the real distance between two or multiple virtual lines on the road, or the actual size of a road region (see Figure 9). Then, the problem of distance estimation is posed as a detection problem in which all vehicles are detected at the same distances whenever they cross the predefined virtual lines or regions. Since the virtual lines or regions are placed on the road, accurate distance estimation involves the accurate location of the contact point of some part of the vehicle. This part of the vehicle should be
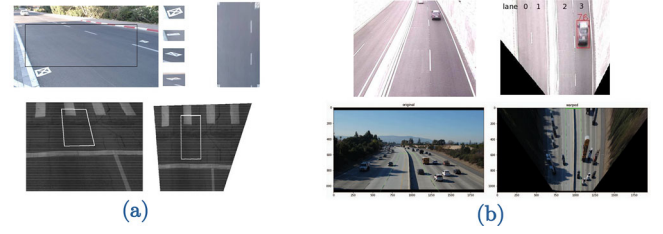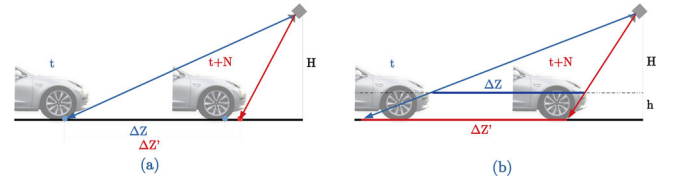
the same at the second location to obtain a coherent estimation of the speed. This is a complex task to be solved due to perspective constraints, spatial (depending on the camera resolution and camera-to-vehicle distance) and temporal (depending on the camera frame rate and the vehicle speed) discretization problems, shadows etc. These problems can be mitigated by using multiple intrusion lines [20, 107].

Second, we have those approaches based on the computation of the homography as a linear transformation of a plane (the road) from the 3D projective space to a 2D projective space (the camera image plane). This way the images can be transformed into a bird's eye view (BEV), which is also known as image warping, in which pixel displacements can be directly transformed to real-world distances. This approach is the most commonly used for both fixed systems [15, 16, 18, 20, 22, 34, 37, 52, 53, 61, 66, 94, 98, 99, 110] and drone-based [25, 30, 57, 62, 64, 69, 96, 104–106, 126]. As depicted in Figure 10, and explained in Section 4.2, the computation of the homography involves the calibration of the intrinsic and extrinsic parameters of the camera and previous knowledge of the real dimensions of some part of the scene. The resulting warped image considers all objects to be on the road plane, which implies deformations of the shape of the vehicles. This is not particularly important, since the points used to calculate the distance generally correspond to the contact points between the vehicles and the road, although this assumption is not entirely accurate due to the perspective of the system and the wrong assumption of motion on the road plane, as depicted in Figure 11. The main advantage

of homography-based approaches compared with the use of intrusion lines is that vehicle detection can be made at any point on the road, and therefore camera frame rate and vehicle speed do not have a negative effect on distance estimation. However, a very precise system calibration is required.

Third, we have the approaches based on the formulation of the estimation of distances from the knowledge of the real dimensions of the objects, including license plates [41, 49, 60, 66, 68, 80, 125] and vehicles [42, 62, 78, 103]. On the one hand, approaches based on license plate dimensions are very sensitive to low meter-pixel ratios and pixel localization errors. On the other hand, such approaches based on vehicle dimensions of the vehicles need to apply fine-grained car model classification systems [135] to identify the specific model and use its specific dimensions.

6.3 | Speed error basics

In a simple scenario, speed is computed using two different locations Z_1 and Z_2 (which represents a linear segment $\Delta Z = |Z_2 - Z_1|$) of a vehicle detected at times t_1 and t_2 ($\Delta t = |t_2 - t_1|$), respectively. The average speed of that vehicle will be estimated as:

$$v = \frac{|Z_2 - Z_1|}{|t_2 - t_1|} = \frac{\Delta Z}{\Delta t}. \quad (6)$$

Taking into account the distance estimation errors at the two locations Z_{1err} and Z_{2err}, in the worst case, we could obtain a speed of:

$$v' = \frac{\Delta Z + Z_{2err} + Z_{1err}}{\Delta t}, \quad (7)$$

resulting in an absolute speed error of:

$$v_{err} = |v' - v| = \frac{|Z_{2err} + Z_{1err}|}{\Delta t} = \frac{|Z_{2err} + Z_{1err}|}{\Delta Z}v, \quad (8)$$

with the relative speed error equal to:

$$\frac{v_{err}}{v} = \frac{|Z_{2err} + Z_{1err}|}{\Delta Z}. \quad (9)$$

We can merge Equations (4) or (5) into Equations (8) and (9) to see that the speed error is proportional to the square of the distance from the camera to the vehicle, which clearly reinforces the requirement that speed measurements should be made with the vehicle as close to the camera as possible. In addition, Equations (8) and (9) reveal the importance of road segment size in the accuracy of speed estimation. That is, the quadratic penalty with the distance is compensated linearly with the length of the road section, which for segments much larger than the distance errors, results in a speed error that grows linearly with the relative distance of the vehicles. This is the main reason why average vehicle speed cameras are so accurate. These systems are based
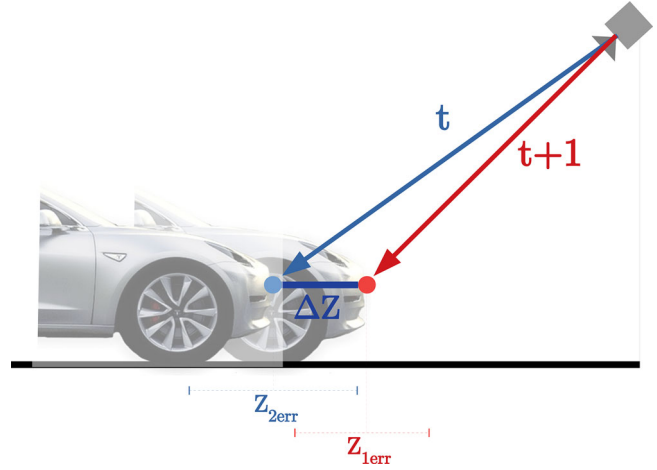

FIGURE 12 Illustration of the speed errors when using consecutive frames. The distance traveled by the detected and tracked features ΔZ (the vehicle) has the same order of magnitude as the distance errors Z_{2err} and Z_{1err}, giving an inaccurate estimate of speed.

on two cameras located several kilometers apart. To meet the requirements of Equation (1), the errors in measuring relative vehicle distances can be several meters and still, maintain a relative speed error within the range of 1–3%.

In the case of fixed speed cameras, the main problem of using the largest possible section of the road is that we would be considering measurements at long distances, with the mentioned penalty in the speed error. Following the procedure proposed in [80], if we differentiate Equation (8) with respect to the relative vehicle distance, an optimal value for the section of the road can be obtained where the speed error is the minimum possible. This optimal stretch of road may fall outside the limits of the image plane when using one camera. That is why a two-camera-based system is proposed in [80] with different focal lengths, similar to the one presented in [61] but pointing at closer distances.

The size of the road section used to calculate vehicle speed is also directly related to the time between measurements. Therefore, when using consecutive images, even with low camera frame rates, the length of the segment ΔZ can be about the same order of magnitude as the distance errors Z_{err}, making it almost impossible to obtain accurate speed measurements (see Figure 12).

Summarizing, we can state the following requirements to minimize the speed error:

- Use distance measurements of the vehicles when they are closer to the camera.
- Avoid using consecutive measurements with short sections.
- When possible, use the optimal road segment to minimize the speed error.
- If the optimal road section falls outside the image plane, use the largest possible road stretch.

Other sources of errors include the violation of the flat-road and constant speed assumptions, which are especially relevant when measuring speed on long stretches of road.

6.4 | Speed computation

6.4.1 | Traffic and individual speed

As described in the proposed taxonomy, the first possible categorization refers to the application domain. On the one hand, we have such approaches focused on measuring the average speed of all vehicles along a road segment, resulting in the traffic speed, or approaches where the accuracy of the measured speed of each individual vehicle is not so important as they are intended to be used as a traffic speed measurement system (with an application context that allows higher margins of error) [12, 17, 20, 23, 32, 34, 38, 40, 42, 51, 53, 54, 58, 61, 69, 77, 79–81, 84, 85, 94, 102, 105, 110, 113, 121]. The rest of the approaches we found are based on obtaining the precise speed of each individual vehicle. Obviously, when these approaches are based on systems that cover all lanes of the road, they can be used to obtain the average speed of the traffic.

6.4.2 | Time/distance between measurements

Once the relative distances of the vehicle, or part of it, have been computed, the speed calculation involves the application of Equation (6). The first requirement is, therefore, related with associating a specific timestamp per each measurement (or image). Prior knowledge about the camera frame rate can be used. However, it is much more accurate and robust to include specific timestamps per each recorded image or measurement taken from the recording system clock (which can be globally and accurately synchronized using NTP servers [80]).

Depending on the camera pose w.r.t. the vehicles and the road, and the speed of the vehicles, a set of N measurements are available per each vehicle. The first decision to be made is which distance between measurements to use to obtain the speed values.

Surprisingly, and contrary to the evidence described in subsection 6.3, most of the works propose to use **consecutive** measurements (at frame t and at frame $t + 1$) to calculate the speed values for each vehicle [14, 18, 19, 22, 25, 33, 35–37, 41, 46, 48, 50, 55, 65, 67, 68, 71–76, 78, 82, 83, 85, 88, 89, 92, 96, 100, 103, 120, 129]. As illustrated in Figure 12, this technique tends to provide noisy values since the distance errors are of the same order of magnitude as the distance traveled by the vehicle. In the case of the use of **non-consecutive** images, we distinguish between several techniques. First, using a fixed distance or region between measurements (e.g. approaches using intrusion lines or regions) [19, 20, 23, 24, 26, 28, 32, 38, 47, 59, 86, 87, 99, 107, 124]. Second, using a predefined number of frames between measurements [15, 52, 93, 108, 111, 125] or a fixed time interval [31, 49, 111] (which are equivalent). Other approaches use the maximum possible distance between the first and last

detection of the tracked vehicle [43, 44, 66, 70, 79, 118, 128]. Finally, we can identify a few techniques that make use of vehicle detection from two different cameras [80] or from different views of a drone camera [106].

6.4.3 | Measurement integration

Once all the N speed values have been computed for a vehicle, the next decision to be made is how to integrate all the measurements to compute the final vehicle speed. Note that all approaches based on virtual intrusion lines or regions [19, 20, 23, 24, 26, 28, 32, 38, 47, 59, 86, 87, 99, 107, 124] are based on a single shot ($N = 1$) so, strictly speaking, no integration of measurements is applied. This approach is commonly referred to as average speed detection.

Most approaches perform instantaneous speed detection, that is, the vehicle speed is calculated and provided as an output for each pair of measurements (using Equation (6)), whether consecutive or non-consecutive [18, 25, 33, 43, 44, 55–57, 60–64, 66, 67, 70–76, 79, 82–84, 88, 93, 95–97, 100, 103, 108, 111, 112, 115–118, 120, 128]. In some cases, the instantaneous speed is filtered using different techniques such as moving average [98], polynomial filtering [129] or using a specific tracker (e.g. Kalman filter) [19, 22, 53, 59, 65, 78, 84, 89, 104].

Another common approach is to compute the mean value for all the available (N) speed measurements [12, 17, 20, 23, 27, 34, 38, 40, 42, 45, 51, 53, 54, 58, 69, 77, 80, 81, 85, 92, 94, 102, 105, 110, 121, 125]. In [80] the mean is computed only using the optimal speed values that correspond to the minimum system speed error.

6.4.4 | Other approaches

Although it is an uncommon approach, we found some methods that use an extended camera shutter to generate motion blur of the vehicles which can be analyzed to obtain the speed of the vehicle from lateral views [13, 21, 39].

In [119] projection histograms are obtained from the result of the frame difference applied in a predefined region. A group of key bins are finally selected to represent the vehicle motion, and all the possible speeds are tested one by one.

Recently we can find a few data-driven approaches based on end-to-end deep learning methods that pose the speed detection problem as a regression problem. For example, in [113] two different convolutional neural networks are trained using animated images (top view of a two lane road segment), from a simulator and synthetic images generated by a cycle-consistent adversarial network (CycleGAN). A standard convolutional architecture is used, with an output fully connected layer of size one (corresponding to the average speed of the road). In [114] vehicle speed estimation is addressed as a video action recognition problem using 3D ConvNets. Finally, a modular neural network (MNN) architecture is presented in [127] to perform vehicle type classification and speed prediction. In all these cases, the input sequence corresponds to a stretch of road with multiple

lanes and vehicles, so they are more suitable for traffic speed detection (individual vehicle speed estimation is only possible when one vehicle is present in the image sequence).

7 | DATASETS AND EVALUATION METRICS

7.1 | Ground truth generation

In order to establish the accuracy of the various approaches to perform vision-based vehicle speed detection, it is necessary to have an appropriate ground truth. As we summarized in Section 3.8, there are a variety of methods and technologies for obtaining the actual speed of the vehicles captured by the cameras. Among these technologies, the most used is the speedometer integrated in the vehicle [31, 35, 38, 42, 47, 79, 83, 84, 93, 96, 104, 106, 110, 112] although this is not the most accurate approach. They are greatly affected by external factors such as the size or pressure of the tires, since the technology consists of directly measuring the rotation speed of the wheel axle, or that of the gearbox axle, and then transforming it into km/h. In addition, the measurements of these speedometers are often biased (overestimated) by the car manufacturer.

Secondly we can find GPS (Global Positioning Systems)-based technologies, which can be categorized into standard-GPS [22, 33, 41, 48, 100, 107], GPS-odometer [65, 66], and differential GPS [58, 64, 80]. The main difference of these technologies lies in the method of obtaining the measurements. Both GPS and DGPS provide global position measurements with which the speed calculation can be computed, taking into account the time difference between measurements. In the case of DGPS some fixed support stations, with known global coordinates, help to correct the measurements returned by the satellites. On the other hand, the GPS odometer directly returns the speed value. Of these three systems, the most precise is the DGPS with an accuracy that can reach a few centimeters.

Other ground truth generation systems that can be found in the literature are range-based sensors such as radar [18, 54, 76, 108, 119] or LiDAR [68, 76, 94], including speed guns, [36, 51, 72, 74, 76, 113] and light barriers [16, 17, 114]. Radar-based sensors use radio waves to determine the range and speed of objects. Radio waves from the transmitter reflect off the vehicle and return to the receiver. The most common and best known type of radar is based on the Doppler effect. On the other hand, LiDAR-based technology is capable of measuring distances by illuminating the vehicles with laser light and measuring the reflection. Differences in laser return times and wavelengths are then used to accurately compute the range. Vehicle speed guns are hand held devices based on radar or laser technology, designed to be light and portable. They can also be mounted on a tripod to avoid hand movements. Light barriers use a beam of light in a transmitter-receiver system which would serve to detect the intrusion of a vehicle into the barrier with an accuracy of less than 10 centimeters. The main drawback of these systems is due to occlusions between vehicles in multi-lane scenarios.

The last type of technology used to generate vehicle speed ground truth is in-pavement sensors, among which we can differentiate between the inductive loop, which consists of a circuit in which an electric current is induced when the metal mass of a vehicle passes by, and the piezoelectric sensors, which generate electricity when they are subjected to an external mechanical force. In both cases the speed calculation is carried out using the current generated in the electric circuit [53, 77, 82, 125, 128].

Finally, a considerable number of works do not carry out a comparison of their vehicle speed detection results with an appropriate ground truth, or do not provide sufficient information about it, which makes it impossible to know the real accuracy of their approaches [12–15, 19–21, 23–28, 34, 37, 39, 40, 43–46, 49, 50, 52, 55, 56, 59–63, 67, 70, 71, 73, 75, 81, 86–89, 95, 99, 105, 111, 117, 118, 120, 121].

7.2 | Available datasets

Although vehicle speed detection is a widely studied topic, the number of datasets available is very limited compared to other research areas and domains. In Table 1 we present an overview of the found datasets and their main features.

The most widely used dataset for traffic and vehicle speed detection in recent years is the AI CITY CHALLENGE [88, 89, 91, 98]. It is part of an annual challenge proposed by NVIDIA, and each year the number of videos and samples provided varies. Specifically, for the year 2018 it consisted of three video tracks. The first one contains 27 one-minute sequences, with a resolution of 1920✕1080 pixels at 30 fps. The second track includes 100 15-minute sequences with a resolution of 800✕410 pixels at 30 fps. The last track contains 15 videos between half an hour and an hour and a half long, at 1080p at 30 fps. All these videos were recorded in normal traffic situations such as highways or intersections with a very high meter-to-pixel ratio, which make them unsuitable for accurate vehicle speed detection. Actually, no speed ground truth is provided.

Another dataset is the BrnoCompSpeed [85, 123], which contains 18 videos at a resolution of 1920✕1080 pixels at 50 fps, including more than 20,000 tagged vehicles. Speed ground truth is provided using a laser-based light barrier system. The sequences were recorded in real traffic situations with a high-medium meter-to-pixel ratio. Appropriate calibration parameters are available.

The dataset available with the lowest meter-to-pixel ratio is the one generated by the Federal University of Technology of Paraná [53, 82, 122] (we refer to this dataset as **UTFPR** dataset). It contains up to 5 h of videos recorded from one camera covering three lanes in different weather conditions, with associated ground truth speeds obtained by inductive loop detectors which are properly calibrated and approved by the Brazilian national metrology agency. The images have a 1920✕1080 pixel resolution and they were captured at 30 fps. The scenario of the dataset corresponds to a urban environment where the maximum speed limit was 60 km/h.

TABLE 1   Overview of the vehilce speed detection datasets

Dataset	Application domain	# Lanes	Meter-pixel-ratio	Resolution and sampling rate	Duration	Speed ground truth
AI CITY CHALLENGE [88, 89, 91, 98]	Traffic	2-8	Very high	1920×1080 @30Hz	40 h	None
BrnoCompSpeed [85, 123]	Speed camera	2	High-medium	1920×1080 @50Hz	20,865 tagged vehicles	Laser-based light barrier
UTFPR dataset [53, 82, 122]	Speed camera	3	Medium-low	1920×1080 @30Hz	5 h	Inductive loop detectors
QMUL junction [103]	Traffic	6	Very high	360×288 @25Hz	1 h	None

Other datasets not specifically devised for vehicle speed detection, are the **QMUL** junction dataset [103] which contains 1 hour of real traffic images at intersections at 25 fps with a resolution of 360×288 and the one compiled by the University of Beihang in China, which contains 70 images of real traffic on highways taken by a UAV, with a resolution of 960×540 [69]. No speed ground truth is available in both cases.

7.3 | Evaluation metrics

To measure the speed error produced by the systems, it is necessary the use of metrics to clearly assess their accuracy. The most commonly used metric is the **Mean Absolute Error** (MAE) either absolute or relative as a percentage [16–18, 31–34, 41, 48–50, 65–68, 70, 72, 74, 79, 80, 83, 85, 92, 99, 102, 104, 106–108, 110–114, 117–120, 124, 125, 129]. The MAE is usually provided with other variables such as the standard deviation, or other statistical metrics, to bound the accuracy of the speed estimation system.

Apart from the MAE, we can find other metrics [14, 15, 19, 24, 77], in which the total error of a mathematical function is calculated, including and modeling the error sources of the speed detection system, such as the pixel detection error, the length measurement error, or the camera calibration error. With these known errors, and some error propagation techniques, such as uncertainty propagation using partial derivatives, the total system error can be accurately obtained.

Finally, the use of the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) have been proposed in several works [38, 40, 64, 74 82, 88–92, 98, 100–102 105, 113, 114, 116, 129]. The MSE/RMSE measures the "distance" between the estimates and the real value of what is being measured, using a sum of square differences between these two values. In some works this method is used to measure the relative distance error, finally measuring the speed error as the absolute difference between the real values (ground truth) and the computed ones.

8 | DISCUSSION

As illustrated in the Figure 4, there has been a significant increase in the number of proposals that deal with vehicle speed estimation using computer vision techniques. This can be explained by the combination of two factors: the increasing importance of accurate and effective vehicle speed detection for speed limit enforcement and traffic monitoring to increase road safety and improve traffic conditions, and recent advances in the computer vision field. However, after the in-depth analysis carried out in this paper, we can state that the set of computer vision approaches available to solve this particular problem is still somewhat behind than in other application contexts, leaving a long way to go. This is especially applicable in the case of speed enforcement due to the challenge of strict requirements on precision and robustness.

In what follows, the state of the art is summarized and discussed, focusing our analysis on four main points: lighting/weather conditions, benchmarking, best practices and data driven challenges.

8.1 | Lighting/weather conditions

Except for some exceptions [48], the detection of vehicle speed has been addressed in good lighting (daytime) and weather (mostly sunny) conditions. This shows that the level of maturity of the subject is still insufficient. Vehicle speed detection should be robust to challenging weather conditions, including rainy, foggy and snowy, since speed enforcement and traffic management are of great importance in such scenarios. Weather conditions can be recognized [141], the impact of bad weather on images can be removed or mitigated [142], and vehicle detection and tracking approaches can be adapted to deal with complex weather conditions [143].

Dealing with nighttime scenarios is also very challenging due to poor visibility conditions, blurring, and glare from vehicle headlights. Specific nighttime vehicle detection approaches can

be applied [144, 145] to deal with vehicle speed detection. In addition, the use of infrared lighting can be very helpful, especially for those approaches based on the detection and tracking of license plates, since it makes their detection and segmentation easier.

We can therefore infer that accurate vision-based vehicle speed detection in challenging lighting and weather scenarios is a solid line of future research.

8.2 | Benchmarking

In order to evaluate and compare the accuracy of vehicle speed estimation approaches, appropriate public datasets are needed, including, at least, different testing scenarios (with different meter-pixel ratios), a wide variety of types of vehicles, and a wide range of speeds (including over-speeding cases). More advanced features would include multiple lighting (daytime/nighttime) and weather (sunny, cloudy, rainy, foggy and snowy) conditions.

As it has been described, and despite the high number of published papers, the number of available datasets around this topic is very limited. On the one hand, this shows a still low level of maturity of the subject. On the other hand, this can be also explained due to the high complexity and cost of the data recoding and labeling processes. In fact, apart from the conditions mentioned above, the datasets must also contain, for each vehicle, an appropriate labeling, including the location of the vehicles or any of their features or regions for all the images from the sequence, and a speed value (ground truth) obtained from some precision system (e.g. range-based sensors, DGPS, inductive loop detectors etc.) whose measurements must be properly synchronized with the images captured by the cameras.

Datasets must be generated following the best practices described in Section 8.3. Appropriate system calibration parameters (intrinsic and extrinsic) should be also provided for each testing scenario. Other useful variables to be labeled can be the vehicle type, or even the make and model.

Benchmarking also involves a good definition of accuracy metrics. Besides using MAE or MSE/RMSE metrics, it is also important to decide about the metric formulation, that is, instantaneous or averaged from a set of consecutive or non-consecutive frames, or single shot from a fixed region. This formulation will strongly depend on the methodology used to obtain the ground truth values and there may be limitations in measuring the performance of certain approaches. For example, if we use light barriers in two different locations, instantaneous speed detection approaches could not be properly evaluated.

The difficulties mentioned should not discourage researchers since the progress of vision-based vehicle speed detection depends largely on the availability of public datasets with the features described above. Future research projects related to this problem should allocate sufficient resources to create these datasets.

8.3 | Best practices

The evidence for the lack of an adequate benchmark can be illustrated by the following statement: we have not found a single work that provides an experimental comparison with multiple approaches. This is one of the reasons why we have not been able to relate the taxonomy of approaches to their performance. At this stage, we can evaluate best practices for future research, but unfortunately we cannot assess the actual performance of the different methods for vehicle speed detection.

Following the fundamentals of vision-based distance and speed estimation errors described in Sections 6.1 and 6.3, a set of requirements have been identified that must be met to address vehicle speed detection as accurately as possible. On the one hand, it is advisable to design a system that provides the lowest possible meter-pixel ratio by using high focal lenses, large baselines for stereo systems, high resolution cameras, and obtaining the distance measurements of the vehicles as close as possible. On the other hand, to minimize the speed error it is important to avoid the use of consecutive frames to calculate speed values, using optimal road segments when possible, or the largest available stretch of the road if the optimal one is not possible with a single camera.

Although there has been significant progress in approaching the vehicle speed detection problem using computer vision techniques, surprisingly, these requirements are not given in most of the papers reviewed. However, we believe that future approaches should take them into account in order to develop accurate and robust vision-based vehicle speed solutions.

8.4 | Data driven challenges

The impact that data and learning-based approaches have had in other areas has definitely not taken place in the area of vehicle speed detection. Only recently have we found a few works that address this approach, with still preliminary results. The main current limitation is the need for much more data than is currently available. Machine learning-based approaches, and especially deep learning-based ones, require a large amount of data to be able to train complex models capable of solving and generalizing a regression problem from input video sequences.

We can anticipate that a possible solution to this problem is the use of synthetic datasets generated from simulators (e.g. autonomous driving simulators), since they allow almost unlimited samples to be obtained from multiple scenarios, with different camera settings (intrinsic and extrinsic), with absolute control of all the simulation variables involved (types of vehicles, speeds etc.), including different weather and lighting conditions, and without the need for manual labeling. The main challenge will be to minimize the well known sim-to-real gap, that is, the applicability of the models trained with synthetic data to real world scenarios.

Another limitation is based on the format used to provide input stimuli to the models. Currently, the few data-driven proposals found use sequences in which a single vehicle appears

with a static background. The question is how to provide input data for multiple vehicles in multiple lanes. As it happens when applying video action recognition approaches for the classification of vehicle lane changes [146], it will be necessary to incorporate some type of process to generate regions of interests (ROIs) with sufficient spatial and temporal information to be able to learn to distinguish different types of speeds for different types of vehicles and scenarios. This problem can also be mitigated by using a system configuration with a very low meter-pixel ratio focusing on a single road lane.

It is expected that the number of learning-based proposals dealing with vehicle speed detection will grow in the coming years.

9 | CONCLUDING REMARKS

In this paper, a comprehensive analysis of the vision-based vehicle speed estimation problem is presented. We review the available literature in the most important application domains: traffic monitoring and control (traffic cameras) and speed limit enforcement (speed cameras). A new taxonomy is proposed that is based on the main components of a vision-based speed detection system, including camera settings and calibration, vehicle detection and tracking, distance and speed estimation, application domains, and ground truth generation. We provide a detailed review of the most relevant work dealing with vehicle speed estimation using computer vision. Finally, we discuss the current limitations of the state of the art, and outline the most important and necessary future steps for the research community dealing with this challenging problem.

The most important application domains are speed limit enforcement, traffic monitoring and control, and, more recently, autonomous vehicles.

ORCID

David Fernández Llorca https://orcid.org/0000-0003-2433-7110

REFERENCES

1. OIML: Radar equipment for the measurement of the speed of vehicles. International Recommendation. OIML R91 (1990)
2. de Jong, G., et al.: Uncertainty in traffic forecasts: literature review and new results for The Netherlands. Transportation 34, 375–395 (2007)
3. Gates, T.J., et al.: Comparison of portable speed measurement devices. Transp. Res. Rec. 1870, 139–146 (2004)
4. Buch, N., et al.: A review of computer vision techniques for the analysis of urban traffic. IEEE Trans. Intell. Transp. Syst. 12(3), 920–939 (2011)
5. WHO: Speed management. A road safety manual for decision-makers and practitioners. Global Road Safety Partnership (2008).
6. ERSO, EC: Speed Enforcement 2018. European Road Safety Observatory (2018)
7. Wilson, C., et al.: Speed cameras for the prevention of road traffic injuries and deaths. Cochrane Database Syst. Rev. CD004607 (2010)
8. Elvik, R.: Developing an accident modification function for speed enforcement. Safety Sci. 49, 920–925 (2011)
9. Sotelo, M.A., et al.: Vision-based adaptive cruise control for intelligent road vehicles. In: Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS), pp. 64–69.IEEE, Piscataway (2004)
10. Milanés, V., et al.: Vision-based active safety system for automatic stopping. Exp. Syst. Appl. 39(12), 11234–11242 (2012)
11. Milanés, V., et al.: Intelligent automatic overtaking system using vision for vehicle detection. Exp. Syst. Appl. 39(3), 3362–3373 (2012)
12. Schoepflin, T.N., Dailey, D.J.: Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. IEEE Trans. Int. Transp. Syst. 4(2), 90–98 (2003)
13. Lin, H-Y., Li, K-J.: Motion blur removal and its application to vehicle speed detection. In: International Conference on Image Processing. IEEE, Piscataway (2004)
14. Cathey, F.W., Dailey, D.J.: A novel technique to dynamically measure vehicle speed using uncalibrated roadway cameras. In: IEEE Intelligent Vehicles Symposium. IEEE, Piscataway (2005)
15. Grammatikopoulos, L., et al.: Automatic estimation of vehicle speed from uncalibrated video sequences. Proc. Int. Symp. Modern Tech., Edu. Prof. Prac. Geo. Rel. Fields (2005)
16. Litzenberger, M., et al.: Estimation of vehicle speed based on asynchronous data from a silicon retina optical sensor. In: Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, Piscataway (2006)
17. Bauer, D., et al.: Embedded vehicle speed estimation system using an asynchronous temporal contrast vision sensor. EURASIP J. Embedded Syst. 82174, 1–12 (2007)
18. He, X.C., Yung, N.H.C.: A novel algorithm for estimating vehicle speed from two consecutive images. In: Proceedings of IEEE Workshop on Applications of Computer Vision (WACV07). IEEE, Piscataway (2007)
19. Alefs, B., Schreiber, D.: Accurate speed measurement from vehicle trajectories using AdaBoost detection and robust template tracking. Proceedings of IEEE International Transportation Systems Conference (ITSC). IEEE, Piscataway (2007)
20. Zhiwei, H., et al.: Models of vehicle speeds measurement with a single camera. In: Proceedings of International Conference on Computation Intelligence and Security Workshops. IEEE, Piscataway (2007)
21. Lin, H-Y., et al.: Vehicle speed detection from a single motion blurred image. Image and Vision Comp. 26, 1327–1337 (2008)
22. Maduro, C., et al.: Estimation of vehicle velocity and traffic intensity using rectified images. In: Proceedings of International Conference on Image Processing. IEEE, Piscataway (2008)
23. Peng, F., et al.: Camera calibration and near-view vehicle speed estimation. Proc. SPIE Image Proc.: Mach. Vis. App. 6813 (2008)
24. Li, Y., et al.: Vehicle speed measurement based on video images. Proc. Int. Conf. Inn. Comp. Inf. and Cont. (2008)
25. Yamazaki, F., et al.: Vehicle extraction and speed detection from digital aerial images. Proc. of IEEE Int. Geo. and Rem. Sen. Symp. IEEE, Piscataway (2008)
26. Pornpanomchai, C., Kongkittisan, K.: Vehicle speed detection system. Proc. of IEEE Int. Conf. on Sig. and Image Proc. App. IEEE, Piscataway (2009)
27. Wu, J., et al.: An algorithm for automatic vehicle speed detection using video camera. In: International Conference on Computer Science and Education. IEEE, Piscataway (2009)
28. Jing-zhong, W., Xiaoqing, X.: A real-time detection of vehicle's speed based on vision principle and differential detection. In: IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics. IEEE, Piscataway (2009)
29. Mehrubeoglu, M., McLauchlan, L.: Determination of vehicle speed in traffic video. In: Proceedings of SPIE Real-Time Image and Video Processing, vol. 7244. SPIE, Washington DC (2009)

30. Liu, W., Yamazaki, F.: Speed detection of moving vehicles from one scene of QuickBird images. In: Proceedings of Joint Urban Remote Sensing Event. IEEE, Piscataway (2009)

31. Tian, J., et al.: Vehicle Speed Measurement Based on Images Energy. In: Proceedings of Asia-Pacific Conference on Information Processing. IEEE, Piscataway (2009)

32. Celik, T., Kusetogullari, H.: Solar-powered automated road surveillance system for speed violation detection. IEEE Trans. on Ind. Elec. 57(9), 3216–3227 (2010)

33. Dogan, S., et al.: Real time speed estimation of moving vehicles from side view images from an uncalibrated video camera. Sensors 10, 4805–4824 (2010)

34. Rahim, H.A., et al.: Vehicle velocity estimation for traffic surveillance system. Int. J. Comp., Elec., Aut., Cont. Inf. Eng. 4(9), 1465–1468 (2010)

35. Rad, A.G., et al.: Vehicle speed detection in video image sequences using CVS method. Int. Journal Phy. Sci. 5(17), 2555–2563 (2010)

36. Rahim, H.A., et al.: Vehicle speed detection using frame differencing for smart surveillance system. In: Proceedings of the International Conference on Information Science, Signal Processing and their Applications. IEEE, Piscataway (2010)

37. Tan, H., et al.: Vehicle speed measurement for accident scene investigation. In: IEEE International Conference on e-Business Engineering. IEEE, Piscataway (2010)

38. Yan, Y., et al.: Research on vehicle speed measurement by video image based on Tsai's two stage method. In: Proceedings of the International Conference on Computer Science & Education. IEEE, Piscataway (2010)

39. Mohammadi, J., et al.: Vehicle speed estimation based on the image motion blur using RADON transform. In: Proceedings of the International Conference on Signal Processing Systems. IEEE, Piscataway (2010)

40. Madasu, V.K., Hanmandlu, M.: Estimation of vehicle speed by motion tracking on image sequences. In: Proceedings of IEEE Intelligent Vehicles Symposium. IEEE, Piscataway (2010)

41. Czajewski, W., Iwanowski, M.: Vision-based vehicle speed measurement method. In: Proceedings of the International Conference on Computer Vision and Graphics, pp. 308–315. IEEE, Piscataway (2010)

42. Nguyen, T.T., et al.: Compensating background for noise due to camera vibration in uncalibrated-camera-based vehicle speed measurement system. IEEE Trans. on Veh. Tech. 60(1), 30–43 (2011)

43. Ibrahim, O., et al.: Towards speed detection camera system for a RADAR alternative. In: Proceedings of the International Conference on ITS Communications. IEEE, Piscataway (2011)

44. Ibrahim, O., et al.: Speed detection camera system using image processing techniques on video streams. Int. J. Comp. Elect. Eng. 3(6), 711–778 (2011)

45. Liang, W., Junfang, S.: The speed detection algorithm based on video sequences. In: Proceedings of International Conference on Computer Science and Service System. IEEE, Piscataway (2012)

46. Ranjit, S.S.S., et al.: Real-time vehicle speed detection algorithm using motion vector technique. In: Proceedings of the International Conference on Advances in Electrical & Electronics. IEEE, Piscataway (2012)

47. Shen, I.Z, et al.: Vehicle speed detection based on video at urban intersection. Research Journal of Applied Sciences, Eng. and Tech. 5(17), 4336–4342 (2013)

48. Sina, I., et al.: Vehicle counting and speed measurement using headlight detection. In: Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, Piscataway (2013)

49. Garg, M., Goel, S.: Real-time license plate recognition and speed estimation from video sequences. TSI Trans. Electrical Electron. Eng. 1(5), 1–4 (2013)

50. Dehghani, A., Pourmohammad, A.: Single camera vehicles speed measurement. In: Proceedings of the Iranian Conference on Machine Vision and Image Processing (MVIP). IEEE, Piscataway (2013)

51. Li, S., et al.: Video-based traffic data collection system for multiple vehicle types. IET Intell. Transp. Sys. 8(2), 164–174 (2013)

52. Shukla, D., Patel, E.: Speed determination of moving vehicles using Lucas-Kanade algorithm. Int. J. Comp. App. Tech. and Res. 2(1), 32–36 (2013)

53. Luvizon, D.C., et al.: Vehicle speed estimation by license plate detection and tracking. In: Proceedings of the IEEE International Conference on Acoustics, Speech and, Signal Processing. IEEE, Piscataway (2014)

54. Lan, J., et al.: Vehicle speed measurement based on gray constraint optical flow algorithm. Optik 125, 289–295 (2014)

55. Kumar, K.V.K., et al.: Vehicle speed detection using corner detection. In: Proceedings of the International Conference on Signal and Image Processing. IEEE, Piscataway (2014)

56. Gupta, P., et al.: Estimating speed of vehicle using centroid method in MATLAB. Int. J. Comp. App. 102(14), 1–8 (2014)

57. Xin, Z., et al.: Algorithm of vehicle speed detection in unmanned aerial vehicle videos. Proceedings of the Transportation Research Board 93rd Annual Meeting. Transportation Research Board, Washington DC (2014)

58. Qimin, X., et al.: A methodology of vehicle speed estimation based on optical flow. Proc. of IEEE Int. Conf. on Serv. Op. and Log., and Inf. IEEE, Piscataway (2014)

59. Bandara, Y.M.P.C.Y., et al.: Frame feature tracking for speed estimation. In: International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, Piscataway (2014)

60. Lee, I., et al.: The vehicle speed detection based on three-dimensional information using two cameras. In: Proceedings of Ubiquitous Information Technologies and Applications, pp. 229–236. IEEE, Piscataway (2014)

61. Lin, L., et al.: Biologically inspired composite vision system for multiple depth-of-field vehicle tracking and speed detection. In: Proceedings of Asian Conference on Computer Vision (ACCV), pp. 473–486. Springer, Berlin Heidelberg (2014)

62. Moranduzzo, T., Melgani, F.: Car speed estimation method for UAV images. Proceedings of IEEE Geoscience and Remote Sensing Symposium. IEEE, Piscataway (2014)

63. Khan, A., et al.: Speed estimation of vehicle in intelligent traffic surveillance system using video image processing. Int. J. Sci. & Eng. Res. 5(12), 1384 (2014)

64. Salvo, G., et al.: Comparison between vehicle speed profiles acquired by differential GPS and UAV. Paper presented at the meeting of the Euro working group on transportation, Sevilla, 2–4 July 2014

65. Rao, Y.G.A., et al.: Real-time speed estimation of vehicles from uncalibrated view-independent traffic cameras. In: Proceedings of IEEE Region 10 Conference (TENCON). IEEE, Piscataway (2015)

66. Ginzburg, C., et al.: A cheap system for vehicle speed detection. arXiv:1501.06751 (2015)

67. Do, V., et al.: A simple camera calibration method for vehicle velocity estimation. In: International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE, Piscataway (2015)

68. Wu, W., et al.: Vehicle speed estimation using a monocular camera. Proceedings of SPIE 9407, Video Surveillance and Transportation Imaging Applications. SPIE, Washington DC (2015)

69. Ke, R., et al.: Motion-vector clustering for traffic speed detection from UAV video. In: Proceedings of IEEE International Smart Cities Conference (ISC2). IEEE, Piscataway (2015)

70. Makarov, A., et al.: Real-time vehicle speed estimation based on license plate tracking in monocular video sequences. Sensors Transducers J. 197(2), 78–86 (2016)

71. Jeyabharathi, D., Dejey, D.: Vehicle tracking and speed measurement system (VTSM) based on novel feature descriptor: diagonal hexadecimal pattern (DHP). J. Vis. Comm. Image Rep. 40, 816–830 (2016)

72. Sundoro, H.S., Harjoko, A.: Vehicle counting and vehicle speed measurement based on video processing. J. Theor. Appl. Inf. Tech. 84(2), 233–241 (2016)

73. Wang, J.: Research of vehicle speed detection algorithm in video surveillance. In: Proceedings of International Conference on Audio, Language and Image Processing (ICALIP), pp. 349–352. IEEE, Piscataway (2016)

74. Jalalat, M., et al.: Vehicle detection and speed estimation using cascade classifier and sub-pixel stereo matching. In: International Conference on Signal Processing and Information Security (ICSPIS). IEEE, Piscataway (2016)

75. Yabo, A., et al.: Vehicle classification and speed estimation using Computer Vision techniques. In: Proc. of AADECA. IEEE, Piscataway (2016)

76. Eslami, H., et al.: Precise vehicle speed measurement based on a hierarchical homographic transform estimation for law enforcement applications. IEICE Trans. Inf. Sys. E99.D(6), 1635–1644 (2016)

77. Filipiak, P., et al.: NSGA-II based auto-calibration of automatic number plate recognition camera for vehicle speed measurement. Proc. European Conf.on the App. of Evolutionary Computation, pp. 803–818 (2016)

78. Han, J., et al.: Vehicle distance estimation using a mono-camera for FCW/AEB systems. Int. Journal of Automotive Tech. 17(3), 483–491 (2016)

79. Enjat Munajat, M.D., et al.: Vehicle detection and tracking based on corner and lines adjacent detection features. In: International Conference on Science in Information Technology (ICSITech), pp. 244–249.IEEE, Piscataway (2016)

80. Llorca, D.F., et al.: Two-camera based accurate vehicle speed measurement using average speed at a fixed point. Proc. of IEEE Intell. Transp. Sys. Conf. (ITSC), pp. 2533–2538.IEEE, Piscataway (2016)

81. Kumar, T., Kushwaha, D.S.: An efficient approach for detection and speed estimation of moving vehicles. Proc. of Int. Conf. Of Inf. Proc. (IMCIP), pp. 726–731.IEEE, Piscataway (2016)

82. Luvizon, D.C., et al.: A video-based system for vehicle speed measurement in urban roadways. IEEE Trans. on Intell. Transp. Sys. 18(6), 1393–1404 (2017)

83. Wicaksono, D., Setiyono, B.: Speed estimation on moving vehicle based on digital image processing. Int. Journal of Comp. Sci. and Applied Math. 3(1), 21–26 (2017)

84. Gerat, J., et al.: Vehicle speed detection from camera stream using image processing methods. In: Proceedings of International Symposium. ELMAR. Croatian Society Electronics in Marine, Zadar (2017)

85. Sochor, J., et al.: Traffic surveillance camera calibration by 3D model bounding box alignment for accurate vehicle speed measurement. Comp. Vis. Image Understanding 161, 87–98 (2017)

86. Setiyono, B., et al.: Vehicle speed detection based on gaussian mixture model using sequential of images. J. Phys.: Conf. Series (ICoAIMS) 890, 012144 (2017)

87. Czapla, Z.: Vehicle speed estimation with the use of gradient-based image conversion into binary form. Proc. Sig. Proc. Alg. Arch. Arrang. Apps. (SPA), pp. 213–216 (2017)

88. Giannakeris, P., et al.: Speed estimation and abnormality detection from surveillance cameras. Proc. IEEE Comp. Vis. Mach. Intell., pp. 93–99 (2018)

89. Kumar, A., et al.: A semi-automatic 2D solution for vehicle speed estimation from monocular videos. Proc. IEEE Comp. Vis. Mach. Intell., pp. 137–144 (2018)

90. Bhardwaj, R., et al.: AutoCalib: automatic traffic camera calibration at scale. ACM Trans. Sen. Netw. 19, 1–27 (2018)

91. Hua, S., et al.: Vehicle tracking and speed estimation from traffic videos. Proc. IEEE Comp. Vis. Patt. Rec. (CVPR) (2018)

92. Bouziady, A.E., et al.: Vehicle speed estimation using extracted SURF features from stereo images. Proc. IEEE Int. Conf. Intell. Sys. and Comp. Vis. (ISCV) (2018)

93. Kurniawan, A., et al.: Speed monitoring for multiple vehicle using closed circuit television (CCTV) camera. Proc. Int. Conf. Comp. Eng., Net. Intell. Mult. (CENIM) (2018)

94. Famouri, M., et al.: A novel motion plane-based approach to vehicle speed estimation. IEEE Trans. Intell. Transp. Sys. 20(4), 1237–1246 (2018)

95. Koyuncu, H., Koyuncu, B.: Vehicle Speed detection by using Camera and image processing software. The Int. Journal Eng. Sci. (IJES) 7(9), 64–72 (2018)

96. Long, H., et al.: Automatic vehicle speed estimation method for unmanned aerial vehicle images. Journal Inf. Hid. Mult. Sig. Proc. 9(2), 442–451 (2018)

97. Patel, N.S., et al.: An algorithm for speed estimation and speed violation detection of a vehicle using computer vision. Int. Journal Adv. Res. Comp. Sci. 9(2), 504–507 (2018)

98. Huang, T.: Traffic speed estimation from surveillance video data. Proc. IEEE Comp. Vis. Patt. Rec. (CVPR), pp. 161–165 (2018)

99. Kim, J-H., et al.: Reliability verification of vehicle speed estimate method in forensic videos. Foren. Sci. Int. 287, 195–206 (2018)

100. Yang, L., et al.: Vehicle speed measurement based on binocular stereovision system. IEEE Access 7, 106628–106641 (2019)

101. Bartl, V., Herout, A.: OptInOpt: dual optimization for automatic camera calibration by multi-target observations. Proc. of IEEE Int. Conf. on Adv. Video and Signal Based Surv. (AVSS) (2019)

102. Tourani, A., et al.: Motion-based vehicle speed measurement for intelligent transportation systems. Int. Journal of Image, Graph. and Sig. Proc. 11, 42–54 (2019)

103. Moazzam, M., et al.: Image-based vehicle speed estimation. Journal of Comp. and Comm. 7, 1–5 (2019)

104. Liu, Y., et al.: Multiple objects tracking based vehicle speed analysis with Gaussian filter from drone video. Intell. Sci. and Big Data Eng. Vis. Data Eng. (IScIDE), pp. 362–373 (2019)

105. Li, J., et al.: An adaptive framework for multi-vehicle ground speed estimation in airborne videos. Remote Sens. 11, 1241 (2019)

106. Biswas, D., et al.: Speed estimation of multiple moving objects from a moving UAV platform. ISPRS Int. J. Geo-Inf 8, 259 (2019)

107. Javadi, S., et al.: Vehicle speed measurement model for video-based systems. Comp. & Elec. Eng. 76, 238–248 (2019)

108. Sochor, J., et al.: Comprehensive data set for automatic single camera visual speed measurement. IEEE Trans. on Intell. Transp. Sys. 20(5), 1633–1643 (2019)

109. Wang, C., Musaev, A.: Preliminary research on vehicle speed detection using traffic cameras. Proc. IEEE Int. Conf. Big Data (2019)

110. Gunawan, A.A.S., et al.: Detection of vehicle position and speed using camera calibration and image projection methods. Procedia Comp. Sci. 157, 255–265 (2019)

111. Julina, J.K.J., et al.: Vehicle speed detection system using motion vector interpolation. Proc. Global Conf. Adv. Tech. (GCAT) (2019)

112. Jiang, J., et al.: Study on a real-time vehicle speed measuring method at highway toll station. Proc. Int. Conf. Sen. Inst. IoT Era (ISSI) (2019)

113. Lee, J., et al.: Image-based learning to measure the space mean speed on a stretch of road without the need to tag images with labels. Sensors 19, 1227 (2019)

114. Dong, H., et al.: Vehicle speed estimation based on 3D ConvNets and non-local blocks. Future Internet 11, 123 (2019)

115. Gauttam, H.K., Mohapatra, R.K.: Speed prediction of fast approaching vehicle using moving camera. Proc. Int. Conf. Comp. Vis. Ima. Proc. (CVIP), pp. 423–431 (2019)

116. Khan, M., et al.: Multiple moving vehicle speed estimation using Blob analysis. Proc. World Conf. Inf. Sys. Tech. (WorldCIST), pp. 303–314 (2019)

117. Murashov, I., Stroganov, Y.: Method of determining vehicle speed according to video stream data. Journal of Physics: Conf. Ser. 1419, 1–7 (2019)

118. Afifah, F., et al.: Vehicle speed estimation using image processing. Journal Adv. Res. Appl. Mech. 48(1), 9–16 (2019)

119. Lu, S., et al.: A high accurate vehicle speed estimation method. Soft Comp. 4, 1283–1291 (2020)

120. Kamoji, S., et al.: Image processing based vehicle identification and speed measurement. Proc. of Int. Conf. on Inventive Comp. Tech. (ICICT) 4, 523–527 (2020)

121. Cheng, G., et al.: Real-time detection of vehicle speed based on video image. Proc. of Int. Conf. on Meas. Tech. and Mech. Aut., pp. 313–317 (2020)

122. Sonth, A., et al.: Vehicle speed determination and license plate localization from monocular video streams. Proc. of Int. Conf. on Comp. Vis. and Image Proc., pp. 267–277 (2020)

123. Mini, T.V., Vijayakumar, V.: Speed estimation and detection of moving vehicles based on probabilistic principal component analysis and new digital image processing approach. Proc. of EAI Int. Conf. on Big Data Innov. for Sust. Cogn. Comp., pp. 221–230 (2020)

124. Dahl, M., Javadi, S.: Analytical modeling for a video-based vehicle speed measurement framework. Sensors 20, 160 (2020)

125. Vakili, E., et al.: Single-camera vehicle speed measurement using the geometry of the imaging system. Mult. Tools Apps. 79, 19307–19327 (2020)

126. da Silva Bastos, M.E., et al.: Vehicle speed detection and safety distance estimation using aerial images of brazilian highways. Proc. Seminário Integrado Soft. Hard. (2020)

127. Madhan, E.S., et al.: A novel approach for vehicle type classification and speed prediction using deep learning. Journal Comp. Theor. Nano. 17(5), 2237–2242 (2020)

128. Liu, C., et al.: A vision-based pipeline for vehicle counting, speed estimation, and classification. IEEE Trans. on Intell. Transp. Sys., Early Access, pp. 1–14 (2020)

129. Bell, D., et al.: Accurate vehicle speed estimation from monocular camera footage. Annals Phot., Rem. Sen. Spat. Inf. Sci. (ISPRS), pp. 419–426 (2020)

130. Izquierdo, R., et al.: The PREVENTION dataset: a novel benchmark for PREdiction of VEhicles iNTentIONs. Proc. of IEEE Intell. Transp. Sys. Conf.(ITSC), pp. 3144–3121 (2019)

131. Stein, G.P., et al.: Vision-based ACC with a single camera: bounds on range and range rate accuracy. Proc. IEEE Intell. Veh. Sym. (2003)

132. Llorca, D.F., et al.: Error analysis in a stereo vision-based pedestrian detection sensor for collision avoidance applications. Sensors 10, 3741–3758 (2010)

133. Orghidan, R., et al.: Camera calibration using two or three vanishing points. Proc. Fed. Conf. Comp. Sci. Inf. Sys., pp. 123–130 (2012)

134. Álvares, S., et al.: Hierarchical camera auto-calibration for traffic surveillance systems. Expert Systems with Applications 41, 1532–1542 (2014)

135. Corrales, H., et al.: CNNs for fine-grained car model classification. EUROCAST, LNCS, 12014, pp. 104–112 (2020)

136. Corrales, H., et al.: Simple baseline for vehicle pose estimation: experimental validation. IEEE Access 8, 132539–132550 (2020)

137. Fernández Llorca, D., et al.: Assistive intelligent transportation systems: the need for user localization and anonymous disability identification. IEEE Int. Transp. Sys. Mag. 9(2), 25–40 (2017)

138. Fernández, C., et al.: A comparative analysis of decision trees based classifiers for road detection in urban environments. Proc. IEEE Int. Transp. Sys. Conf. pp. 719–724 (2015)

139. Anagnostopoulos, C.E., et al.: License plate recognition from still images and video sequences: a survey. IEEE Trans. Intell. Transp. Sys. 9(3), 377–391 (2008)

140. Du, S., et al.: Automatic license plate recognition (ALPR): a state-of-the-art review. IEEE Trans. Circ. Sys. Vid. Tech. 23(2), 311–325 (2013)

141. Yan, X., et al.: Weather recognition based on images captured by vision system in vehicle. In: Advances in Neural Networks - ISNN 2009, pp. 390–398.Springer, Berlin, Heidelberg (2009)

142. Luo, Y., et al.: Removing rain from a single image via discriminative sparse coding. Proc. IEEE International Conference on Computer Vision (ICCV), pp. 3397–3405.IEEE, Piscataway (2015)

143. Rezaei, M., et al.: Robust vehicle detection and distance estimation under challenging lighting conditions. IEEE Trans. Int. Trans. Sys. 16(5), 2723–2743 (2015)

144. Alcantarilla, P.F., et al.: Night time vehicle detection for driving assistance lightbeam controller. In: Proceedings of IEEE Intelligent Vehicles Symposium. IEEE, Piscataway (2008)

145. Chen, Y-L., et al.: A real-time vision system for nighttime vehicle detection and traffic surveillance. IEEE Trans. Ind. Elec. 58(5), 2030–2044 (2011)

146. Biparva, M., et al.: Video action recognition for lane-change classification and prediction of surrounding vehicles. arXiv:2101.05043 (2021)

147. Posch, C.: Bio-inspired vision. Topical Work. Elec. Part. Phy. 2011, IOP Publishing Ltd and SISSA (2012)

148. Li Zhang, L., et al.: Global data association for multi-object tracking using network flows. In: IEEE Computer Vision and Pattern Recognition. IEEE, Piscataway (2008)