Recent progress in road and lane detection: a survey

Aharon Bar Hillel · Ronen Lerner · Dan Levi · Guy Raz

Abstract The problem of road or lane perception is a crucial enabler for advanced driver assistance systems. As such, it has been an active field of research for the past two decades with considerable progress made in the past few years. The problem was confronted under various scenarios, with different task definitions, leading to usage of diverse sensing modalities and approaches. In this paper we survey the approaches and the algorithmic techniques devised for the various modalities over the last 5 years. We present a generic break down of the problem into its functional building blocks and elaborate the wide range of proposed methods within this scheme. For each functional block, we describe the possible implementations suggested and analyze their underlying assumptions. While impressive advancements were demonstrated at limited scenarios, inspection into the needs of next generation systems reveals significant gaps. We identify these gaps and suggest research directions that may bridge them.

Keywords Lane detection · Road detection · Road segmentation · Advanced driver assistance systems

All authors contributed equally to this work.

A. Bar Hillel (✉) · R. Lerner · D. Levi · G. Raz
Advanced Technical Center Israel, General Motors R&D,
7 HaMada St., 46725 Herzliya, Israel
e-mail: aharon.barhillel@gm.com; aharon.barhillel@gmail.com

R. Lerner
e-mail: ronen.lerner@gm.com

D. Levi
e-mail: dan.levi@gm.com

G. Raz
e-mail: guy.raz@gm.com

1 Introduction

Advanced driver assistance systems, which either alert the driver in dangerous situations or take an active part in the driving, are gradually being incorporated into vehicles. Such systems are expected to grow more and more complex towards full autonomy during the next decade. The main bottleneck in the development of such systems is the perception problem [1], which has two elements: road and lane perception and obstacle (i.e. vehicles and pedestrian) detection. In this survey we consider the first.

Road color and texture, road boundaries, and lane markings are the main perceptual cues for human driving. Semi and fully autonomous vehicles are expected to share the road with human drivers and would therefore most likely continue to rely on the same perceptual cues that humans do. While there could be, in principle, different infrastructure cuing for human drivers and vehicles (e.g. lane marks for humans and some form of vehicle-to-infrastructure communication for vehicles) it is unrealistic to expect the huge investments required to construct and maintain such double infrastructure, with the associated risk in mismatched marking [2]. Road and lane perception via the traditional cues remains therefore the most likely path for autonomous driving.

Road and lane understanding includes detecting the extent of the road, the number and position of lanes, merging, splitting and ending lanes and roads, in urban, rural and highway scenarios. Although much progress has been made in recent years, this type of understanding is beyond the reach of current perceptual systems. There are several sensing modalities used for road and lane understanding, including monocular vision (i.e. one video camera), stereo, LIDAR, vehicle dynamics information obtained from car odometry or inertial measurement unit (IMU) with global positioning information obtained using global positioning system (GPS) and digital

maps. Vision is the most prominent research area in lane and road detection due to the fact that markings are made for human vision, while LIDAR and global positioning are important complements.

This paper provides several insights into this domain. We present an up-to-date survey of approaches and algorithms for road and lane detection in recent years, updating the survey [3] from 2005. We found that important due to the significant progress made since 2005 as part of the research efforts made for the DARPA Grand Challenge (2005) and Urban Challenge (2007) and for commercial driver assistance systems. Unlike [3], we do not limit the survey to vision-based systems and include systems based on all relevant sensing modalities. We identify a generic system architecture and present each system described in light of this architecture, by breaking down each system to its functional elements. We provide a discussion of the scope of the problem and challenges ahead in relation to present and future driver assistance systems. Finally, we identify the remaining gaps both in research and in system-level evaluation and suggest research directions for bridging these gaps.

The structure of this survey is as follows: in the next section we define the scope of the lane and road detection task and relate it to the automotive industry requirements. In Sect. 3 we provide a comprehensive overview of the different sensing modalities used for the task. Section 4 presents the functional modules and techniques used and Sect. 5 discusses the integration of these modules in a full system. Section 6 discusses experimental evaluation methods and Sect. 7 concludes the survey with a summary and suggestions for future research directions.

2 Problem scope

The lane detection problem, at least in its basic setting, does not look like a hard one. In this basic setting, one has to detect only the host lane, and only for a short distance ahead. A relatively simple Hough transform-based algorithm, which does not employ any tracking or image-to-world reasoning solves the problem in roughly 90% of the highway cases [31]. In spite of that, the impression that the problem is easy is misleading, and building a useful system is a large-scale R&D effort. The main reasons for that are significant gaps in research, high reliability demands, and large diversity in case conditions.

Research gaps During the next decade, more and more semi-autonomous features are expected to be added gradually to vehicles, toward full autonomy. Some of these features are enlisted in Table 1, with a list of relevant publications from the past 5 years. It can be easily seen that Lane Departure Warning (LDW), the most basic of these features, has

received the largest portion of research attention. The lane understanding level required for this feature is identification of the host lane alone and to a distance of several dozens of meters ahead. Significant research effort was also devoted to full autonomy, mainly due to the DARPA challenges [2,20–25,27–29,32]). Complex road and lane understanding, appearing in the middle rows of Table 1 are very little understood.

Since full autonomy is the most complex problem, including all others as sub-tasks, one may get the impression that the features in the middle rows of the table are covered by the research on full autonomous vehicles. This, however, is not true when it comes to onboard lane and road understanding. The reason is that the lack of cost constraints, together with the presence of highly accurate map information in the DARPA challenges, led to solutions with very limited forms of onboard road and lane perception.

A typical vehicle in the DARPA challenge carried multiple LIDARs, Radars, a highly sensitive IMU and computing power of a dozen computers [22]. Moreover, an exact digital map of the road network coupled with updated aerial imagery was supplied to the competitors [24]. The combination of the detailed map information with the exact positioning equipment (GPS+IMU) enabled localization of the vehicle w.r.t the map in resolution of approximately 1 m [24]. Since the typical road width, as well as the width of the lanes in the urban challenge is usually around 4–5 m [2], such resolution is nearly enough for the vehicle to navigate itself blindly—without any onboard perception.

Under these circumstances, the role of onboard lane and road perception in the urban challenge is usually limited to localization validation and minor refinements. This is often achieved using one-dimensional LIDARs pointed downwards, used to verify the vehicle's position within the lane/road. Five of the six finishing competitors in the challenge did not use the vision modality at all (the exception is [2]). In the desert challenge, lanes were not present at all, and road perception was usually limited to very near range (10–15 m) as navigation is based mostly on exact global positioning.

In contrast to the global positioning-based solutions developed for the DARPA challenge, lane and road perception for commercial vehicles has to be conducted with affordable sensors, which currently include mostly vision, GPS and certain radar types. These considerations, as well as reliability issues discussed in Sect. 3, imply that lane and road understanding remain a challenge that should be solved by onboard sensing, as it has been treated in most of the literature till now.

Due to the reasons stated above, the research on autonomous driving had very limited contribution to the problems appearing in the middle rows of Table 1, taken as onboard perception problems. These features, which are planned to be a major driving force in the automotive industry in the

Table 1 Current and future expected automotive features, and their lane/road understanding demands

Feature	Description	Lane/road understanding demands	References
Lane Departure Warning (LDW)	Issue warnings for near lane departure events	Host lane, short distance (40–50 m ahead)	[4,3,5–16]
Adaptive Cruise Control (ACC)	Follow the nearest vehicle in the host lane with safe headway distance	Host lane, short distance	[17]
Lane keeping	Return the car to the lane center when un-signaled lane departure occurs	Host lane, short distance, higher reliability than LDW	
Lane centering	Keep the car in the middle of the lane at all times	Host lane, medium distance, high reliability, lane split (non-linear lane topology) identification	[6]
Lane change assist	Autonomous lane change on demand	Multiple lanes, front and rear, large distance (150 m) ahead	[18,19]
Turn assist	Autonomous turn on driver demand or as part of automatic navigation	Multiple lanes, lane semantics (identify turning lanes), non-linear lane and road topology (splits and merges)	
Full autonomous driving for paved roads	Autonomous driving in city and highway	All of the above plus complex road topologies such as junctions/round-abouts/road under construction	[2,20–24]
Full autonomous driving for cross country driving	Autonomous driving in non-paved areas	Full rough road understanding but somewhat easier than paved-road autonomy with respect to lack of lanes, sparser traffic	[25–30]

Papers are assigned to table rows based on their feature of interest, or based on the lane detection task considered if no explicit feature was mentioned

next decade, were very sparsely covered. The main research challenges are understanding of multiple lanes, farther to the front and to the rear of the vehicle, and perception of complex non-linear lane and road topologies. Some examples for these challenges are illustrated in Fig. 3.

High reliability demands In order to be useful, driver assistance systems should reach very low error rates. For a warning system like LDW, the false alarm rate should be very low, as high rates irritate drivers and lead to system rejection. The exact amount of false alarms acceptable by drivers is still a subject for research [14,15], and some available systems report few false alarms per hour [16]. At 15 frames per second, 1 false alarm per hour means one error in 54,000 frames. For closed-loop autonomous driving features, errors should

be even several orders of magnitude lower. Such low error rates are very hard to achieve in vision-based applications. Other advanced applications of computer vision, such as Web-based visual search or surveillance are typically much more forgiving in terms of the required error rates.

Case diversity The last source of difficulty for a lane detection system is the rich set of conditions that has to be taken care of, requiring development of many different algorithms and sub-systems. The main sources for condition diversity are:

– **Lane and road appearance diversity** Lane marks are typically 0.1 m wide with white or yellow color, but many other variants exist: circular reflectors, cat's-eyes, lane marks with special colors and with variable width.
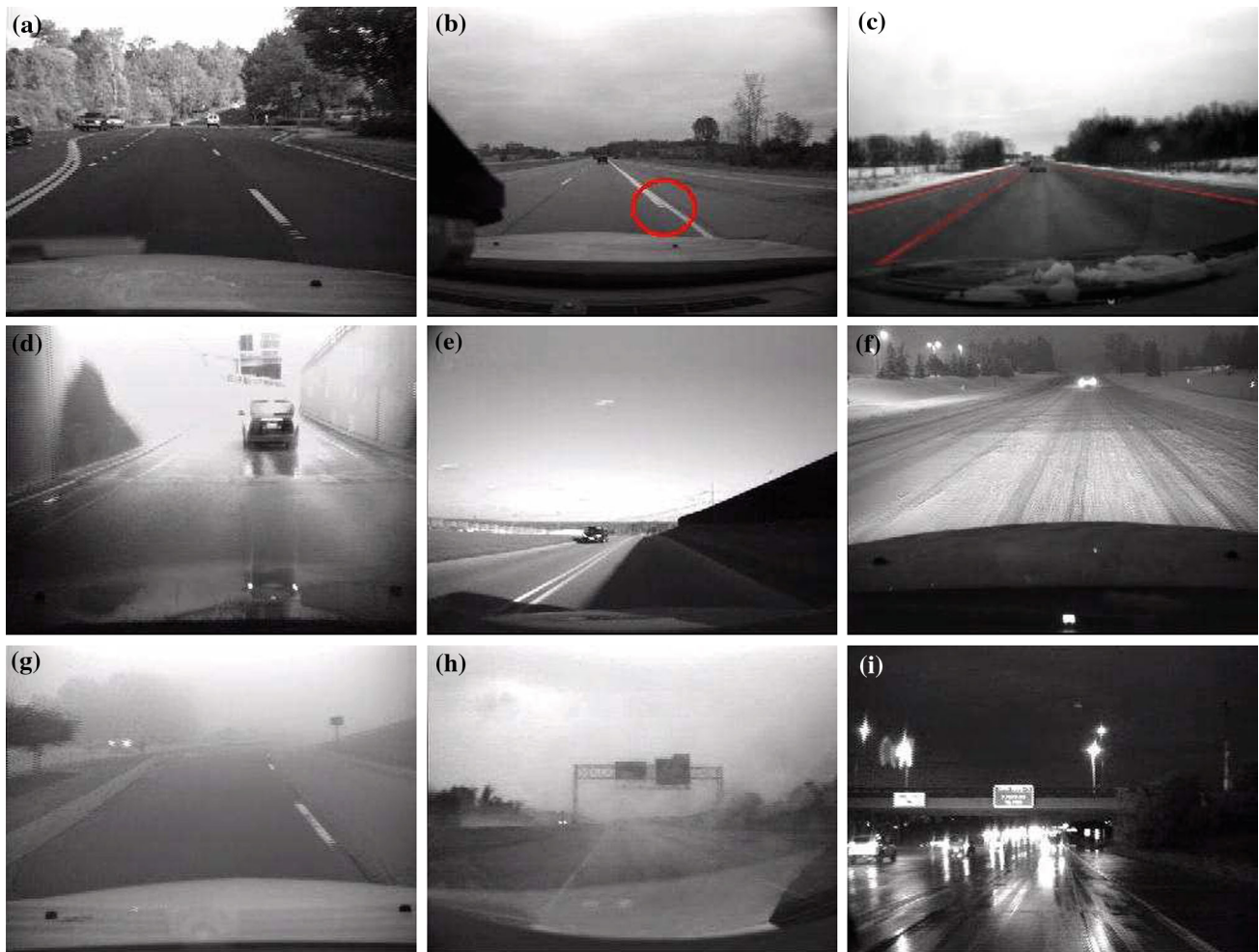
Fig. 1 Scenario diversity which road and lane detection should cope with. Lane and road appearance diversity: **a** different lane marks, **b** the width of the marks are changed (marked by the *circle*), **c** different lane widths (the right lane is extremely wide due to the merge with another lane). Image clarity issues: **d** saturated image's upper part at tunnel exit, **e** cast shadow on the road, **f** road covered with snow. Poor visibility conditions: **g** low visibility due to fog, **h** low visibility due to heavy rain, **i** reflections on wet road at nighttime

Lanes are typically 3.05–3.66 m wide [33], but exceptions (about 12% in the U.S.) are frequent. The number of lanes may vary. The road is usually straight, and curvature is usually limited (for example the curve radius is at least 80 m for 50 KPH urban roads [34]), but exceptions exist.

– **Image clarity issues** While usually the road is open and visible, there are some exceptions which cannot be ignored. Nearby vehicles can create severe occlusions. Shadows from nearby trees and buildings may create misleading edges and texture on the road. In some cases, like when the host vehicle comes out of a tunnel there are abrupt changes of several orders of magnitude in the illumination level, leading to over exposed image.

– **Poor visibility conditions** The system should operate, or at least identify the condition and lower its confidence, under rain, fog, haze and night conditions. Each such condition requires another algorithmic treatment at some processing level.

Figure 1 illustrates some of the challenging scenarios mentioned above. Reasonable coverage for all these conditions requires developing a complex system with considerable engineering effort. Teams which tried to develop vision-based road understanding systems in a fast manner often gave up this modality and turned to other means [22, 23]. Most of the aforementioned difficulties have to be faced even when implementing the most basic features (e.g. LDW).

3 Relevant modalities

As mentioned in the previous section, several different sensors or perception modalities were found useful for the task of lane and road perception. In this section we enlist these modalities and discuss their main characteristics and typical use.

Monocular Vision Monocular vision modality, or more simply put, a camera, is the most frequently used modality for lane and road perception. Most works either discuss this possible modality directly or acknowledge it as a viable alternative. In all available literature, front lane and road detection is considered, so a typical lane detection camera is usually front mounted in the middle of the car. The required resolution can be derived from maximal distance d at which the system is expected to identify a lane mark, by $N_p = Cd/w$ where N_p is the number of horizontal pixels, C is the camera Field of View (FOV) width in radians, and $w = 0.1$ m is the lane mark width. This formula assumes that a lane mark can be identified when its width is a single pixel and gives a requirement of 800 horizontal pixels for lane mark detection at 100 m distance with a 46 degrees camera.

There are two main reasons why vision-based imaging has taken a leading role in the task of lane and road detection: First, visual data are certainly the main modality in use when human drivers are involved. Lane marks and road boundaries are designed so that a human driver will be able to see them in all driving conditions. Since the road and lane mark infrastructure is most suited to the human visual system, using a camera to get the same visual cues for a computational equivalent system makes great sense. In fact, it is generally true that since lane marks have a signature only in the visual domain, no lane detection system could be realized without referring to the visual modality. Second, following the same reasoning from an "evolutionary" viewpoint, it is clear that cameras currently are the cheapest and most robust modality for automotive applications. The maturity of consumer-level mass produced camera modules, coupled with large investment in machine vision allows for good cost-effective solutions to be attainable.

The vision modality can be used essentially in all relevant stages of lane and road processing, as we show in following sections, and much effort is invested into this modality. Still, one has to stress that the robustness of state-of-the-art processing algorithms is still far from satisfactory and much further development is required. Current vision-based algorithm call for the use of many assumptions on road and lane nature and lack in adaptive power compared with the average human driver.

LIDAR Light Detection And Ranging (LIDAR) represents another major possible modality for lane and road detection. As mentioned in the previous section, it has been used to a large extent by teams participating in the DARPA urban challenges. Several vendors provide commercially available LIDARS, and some research institutes are using their own. The major obvious drawback of the LIDAR modality is the relatively high cost of such sensors. The current high cost prevents such sensors from becoming wide-spread commodities for automotive applications.

The LIDAR, being an active Time Of Flight (TOF) device, can measure the 3D structure of the vehicle surrounding. In addition, most LIDARs can report reflected intensity as well, providing a substitute to a visual camera with an advantage of being an active light source and thus independent of natural light issues. This specifically helps in coping with shadows and darkness. Since lane marks have only intensity information and no 3D structure, intensity measurement is required if the LIDAR is to be used as the only modality [2,32,34–36].

The added 3D information supplied by LIDAR has been mostly used in the following tasks:

1. Identification of objects obscuring lane marks and road boundaries by their 3D extension above road surface [37,2].
2. Estimate ground roughness as a basis for road/off-road segmentation [2,32,38]. In turn, this segmentation can lead to road edge detection, off-road cuing, road seeding, etc. [2,22,29,37]
3. Detect curbs and berms as an edge-of-road marks [37,38].

In addition, LIDAR can be used to detect host vehicle pitch and road angles (most notably slopes) in order to improve image to world correspondence (see Sect. 4.5). The use of 3D data instead of 2D image allows for greater robustness and success rates: curbs, berms and road roughness are strong road markers. Obstacles are more easily detected in 3D, as well as road geometry.

We use the term LIDAR to indicate a full 3D measurement apparatus. There is, however, a sub class of one dimensional LIDAR scanners which can produce range measurement over a line cross-section of the road [3,22–24,32]. Such systems can detect road edges and some obstacles but typically lack the look-ahead power.

Stereo imaging Stereo imaging, the use of two cameras in order to obtain 3D information, represents a step between single camera modality and 3D LIDAR. Stereo imaging is typically much cheaper to implement than LIDAR and it inflicts smaller footprint on the host vehicle. On the other hand, stereo imaging generally cannot reach the same range accuracy and reliability that a LIDAR can. Unlike LIDAR, successful depth measurement is texture dependent, with extremely uniform surfaces posing a challenge. The range accuracy is a function of the stereo baseline (the distance between the two cameras). A larger baseline stereo system will provide better range accuracy, but often at the cost of lower reliability and higher computational cost (a tougher correspondence problem to solve). Generally speaking, stereo imaging poses a greater processing challenge compared with LIDAR system, with increased probability of errors.

Still, stereo imaging can be used for the same basic tasks as LIDAR, including obstacle identification, host to road pitch angle estimation [8], curb detection [8,39] and 3D road geometry and slope [8,40] estimation.

Geographic information systems (GIS), GPS, and inertial measurement unit (IMU) The use of prior geographic database together with known host vehicle position has become an everyday activity. Commercial navigation systems are being used worldwide to guide human drivers. As described in Sect. 2, this powerful combination was extensively used in the DARPA urban challenge to guide autonomous vehicles with limited onboard sensing feedback [2,22–24]. Taken one step further, it is conceivable that global positioning with accurate map information will enable worldwide autonomous driving without any onboard lane and road sensing at all (onboard sensing will still be required for obstacle detection though). The degree to which such a vision is realizable depends on accuracy (that is, resolution in meters) and reliability (i.e. failure probability) of both global positioning and map information.

Current commercial GPS receivers typically achieve accuracy of 5–10 m [41], which can be improved to 1 m using accurate IMU integration [22,24]. Possible sources for map information are high-resolution aerial (or satellite) images or GPS measurements collected on the ground [24], converted to digital maps and Digital Elevation Models (DEM). The resolution obtainable from aerial imagery can be 0.25 m and higher [22], so lane marks can be clearly seen in such images. Hence the accuracy gap to driving without sensors is mainly due to the global positioning accuracy.

The more severe obstacle to reliance on global positioning is reliability. GPS success depends on connection with enough satellites, which may be lost due to a variety of reasons. GPS loss-of-contact epochs can partially be tolerated using IMU integration, and the possible reliability of this combination is a current research topic [42,43]. Whether highly accurate digital map information can be obtained and kept updated for large-scale terrains in a reliable manner is most questionable. The bottom line is that commercial road/lane understanding based on exact global positioning cannot be ruled out, but it is not the common assumption in most of the literature and in this survey. Instead, global positioning prior should be considered as an important complementary information.

Vehicle dynamics We refer by vehicle dynamics to the proprioceptive vehicle perception of its speed, yaw rate and acceleration. These are typically measured with internal sensors like wheel speed or stirring angle and reported in the internal vehicle communication bus. This information is often used in the temporal integration module [3,4,9] to enable better tracking of lane/road models and lower-level

features. The accuracy of these measurements is limited, and sometimes an IMU is used for these purposes instead [2].

Radar Radar modality, while useful for other tasks, lacks the resolving power to observe lane marking or even delicate 3D structures. The relevance of RADAR sensors is twofold [44,45]:

1. Detect obstacles (i.e. other vehicles) that obscure the lane marking and road boundaries.
2. Discriminate between road and off-road regions based on their large reflectivity difference.[1]

Both properties form only a limited subset of the capabilities a LIDAR has, although with different related cost and other technical parameters.

Obvious benefits can be gained from the combination of more than one modality. The different modalities can complement the weaknesses of each other and provide higher overall system reliability. Fusion of data from multiple modalities can also provide a way to estimate confidence level by comparing reports of different modalities, an important task at the system level. Another consideration is that certain modalities may be available anyway on the vehicle, due to some other task that requires them, so they can be utilized with very low cost. GPS and Radar are the most prominent examples.

4 Modules and techniques

In this section we break down the road and lane detection task into functional modules and enumerate the possible approaches suggested to the implementation of each module. We start by presenting the system decomposition, including enumeration of the constituent modules in a full generic system and their interconnections. The following subsections present approaches to the implementation of the various modules: image pre-processing (Sect. 4.1), feature extraction (Sect. 4.2), road/lane model fitting (Sect. 4.3), temporal integration (Sect. 4.4), and image to world correspondence (Sect. 4.5).

Inspection of the lane and road detection literature reveals that most of the suggested systems share the main functional modules, though these modules are of course implemented differently in different systems. Based on the commonalities between the algorithms we extracted a generic system for road and lane detection, whose functional decomposition is presented in Fig. 2. The system is generic, as none of

---

[1] A road surface behaves like a mirror with respect to RADAR radiation and thus reflect back very little radiation. Off road surface tend to be more rough and have higher RADAR reflectivity.
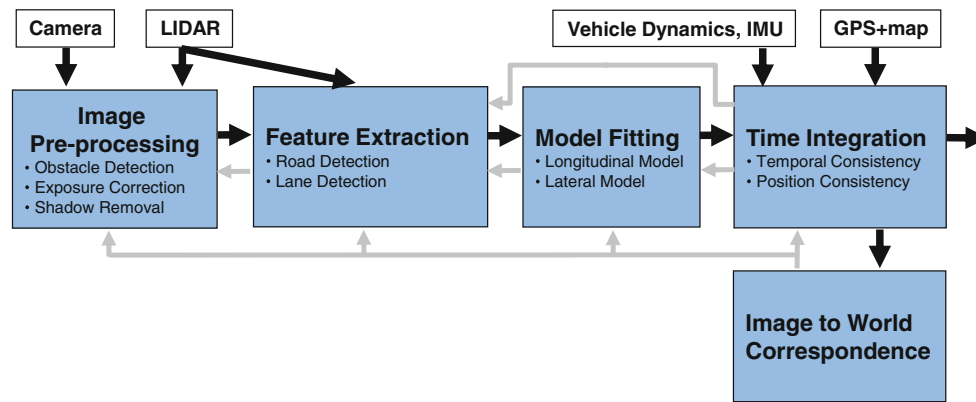
Fig. 2 A generic system: functional decomposition

the systems in the literature has all the modules appearing. However, almost all the algorithms we encountered can be mapped to subsystems of this system, with the most mature systems having nearly all of the modules. Like the flow diagram presented in [3], we use this generic system as a skeleton enabling comparison between different algorithms according to their functional parts.

The main modules we identify, plotted as boxes in Fig. 2 are

- **Image pre-processing** There are several operations which can be applied to the image before feature extraction to reduce clutter and enhance features of interest. Obstacles (mostly vehicles) regions can be identified and removed. Shadows can be significantly weakened using a preprocessing transformation applied to the entire image. Over and under exposure cases can be accounted for by image normalization or by actively controlling the camera exposure. Finally, based on the image to world correspondence, the image area considered can be truncated by removing the region above the horizon or otherwise limiting the region of interest.
- **Feature extraction** Low level features are extracted from the image to support lane and road detection. For road detection, these typically include color and texture statistics allowing road segmentation, road patch classification or curb detection. For lane detection, evidence for lane marks is collected.
- **Road/lane model fitting** A road and lane hypothesis is formed by fitting a road/lane model to the evidence gathered.
- **Temporal integration** The road and lane hypothesis is reconciled with road/lane hypotheses from the previous frame and with global positioning information, if available. The new road/lane hypothesis is accepted if the difference between the new and previous frame can be explained based on the vehicle dynamics.

- **Image to world correspondence** This module provide services of translation between image and ground coordinates, using assumptions about the ground structure and camera parameters. This translation is most required by the temporal integration module, but there are cases in which it is used by all other modules. For example, it may be used to allow features based on consecutive frames substraction [46], or to fit the road model in an inverse perspective image [12].

The main information flow in the system is the bottom up path, indicated by the thick arrows in Fig. 2. However, feedback connections also exist, in which higher level modules guide earlier modules toward better feature extraction or model fitting. Feedback interaction is most common, but not limited to, consecutive stages. Interaction between model fitting and feature extraction is possible in many ways. For example, in [25,26,47] the vanishing point is computed based on texture feature voting, and then lines passing through it are scanned to score possible road edge candidates. In [2] lane candidates are first found using simple filters, and at candidate locations second-order derivatives are computed for further pruning. In [6] a lane model is estimated one image row after the other, starting from the bottom of the image. The lane model gathered from early rows guides the feature extraction at higher ones. Many similar examples exist.

Tight interaction between the temporal integration module and the road model enforcement is also suggested in several papers. For example, in [20] the previously found lane model is transformed into a lane model in the new image, and only the additional possible lane continuations are considered. In [5] the previous lane model constrains the lane directions considered in a Hough transform. Feedback from feature extraction to image pre-processing is always applied when vision-based obstacle detection is used [5]. In addition to tight feedback loops between consecutive stages, the temporal integration module can guide feature extraction in

several ways. In [46] the homomorphy between previous and current frames is used to extract low level features based on image differences and to give higher weight to image regions which were earlier identified as road. In several cases, model lines detected at the previous image define a region of interest in which feature are searched in the new image [4,20]. In [3] line directions in the previously detected model limit the set of oriented filters applied to the image.

An implementation for any of these modules relies on a set of assumptions regarding the camera, the road and the vehicle dynamics. These assumptions may be explicit, for example, in the road and lane model used, or implicit in the choice of algorithmic approach and details. In our description of the possible module implementation we will cluster algorithmic approaches according to the assumptions they use and highlight these assumptions. Such clear presentation of the assumptions allows better understanding of the expected failure modes of certain techniques and of the complementary value of possible technique combinations.

4.1 Image pre-processing

The first functional module in the generic pipeline is image pre-processing. Here, our objective is to remove clutter, misleading imaging artifacts and irrelevant image parts. The remaining cleaned image parts will serve as the input data from which features will later be extracted. In general, methods that fall under this module's scope can be categorized into two families: handling illumination-related effects for enhanced image quality and pruning parts of the image that are suspected as irrelevant for the confronted estimation task.

One aspect of handling illumination-related effects is adapting the dynamic range of the capturing device. A robust road or lane detection system should be capable of coping with different illumination conditions, varying from sunny midday to night time artificial illumination. These changes, although being very large, are also characterized by being very slow and gradual. An abrupt illumination change might also be confronted when entering or exiting a tunnel or while driving below a bridge that casts its shadows on the hosting road. Most systems do not develop their own mechanism for adaptive dynamic range and base their robustness on standard camera capabilities such as automatic iris or gain control, available in many commercial cameras. The adjustments of camera iris or gain are typically performed gradually, which improves the stability of these mechanisms on one hand, but deteriorates their performance in the presence of abrupt illumination changes on the other hand. In [48] four images with different exposures are periodically captured. An adaptive mechanism controls the exact exposure of each of the images, and their fusion yields a very high dynamic range image that can also handle abrupt illumination changes and maintains high-quality image under various scenarios.

Another illumination effect that should not be overlooked is lens flare, caused by direct sunlight in the camera field of view. In [2], date, time and geographical coordinates are used for the computation of a solar ephemeris. Maintaining full camera calibration, then, allows deducing the sun location on the image plane and rejecting straight bright lines pointing at that direction.

Cast shadows on road surface are major source of clutter due to the intensity edges they produce. In order to circumvent this illumination effect some works [5,30,49] perform a variety of color-space transformations to HSL, LAB, YCbCr and others. Then, by combining different color channels illumination invariant images are obtained, where illuminated and shadowed areas of the same surface obtain similar intensity. It should be noted that these techniques are all based on the assumption that hue information still exists at shadowed parts of the surface, implying that there is enough ambient light in the scene. Additionally, they assume that the hue is not biased by the environmental illumination or that this bias can be calibrated and thus compensated. In [30] four illumination-invariant images are proposed. Two of them are obtained by combining different color channels, while the other two are texture based. Edge maps were extracted from the former two images and the edge density was utilized to produce a shadow-free image. It is claimed that these images are invariant to wet versus dry surface appearance. Such texture images are not hue based and thus do not rely on the aforementioned assumptions. They do, however, require that enough texture will be present on road surface. Another possibility to diminish the cast shadows effect is at the feature extraction stage, rather than at the image pre-processing module. In [3] only edges that align with the presumed road boundary directions were extracted, and hence most shadow-related clutter is filtered.

The second category of image pre-processing techniques includes different methods for pruning image parts that are suspected to contain irrelevant or misleading information for the road/lane estimation task. Obstacles, like cars and pedestrians, are major sources for outlying data and different approaches were followed for their detection and removal. In [46] the 2D motion of image points was tracked and structure-from-motion technique was applied to infer if such motions comply with ground plane motion or with off-ground obstacles. Similar steps are followed in [5]. However, the tracking was performed on color-based segmented blobs. Such a technique was also tested by [23] and was found unreliable due to high false-positive rate. In [2,37] 3D data were directly obtained from a LIDAR sensor, enabling simple off-ground point detection and rejection.

Another method to reject irrelevant image parts is by defining Regions Of Interest (ROI) on the image plane. Only these regions will be processed at the feature extraction phase. In its most simplified form, Zhang et al. [47] defines the lower

half of the image as an ROI. In some works like [2,12,27], the connection between the 3D world and the 2D image is estimated using different techniques that are elaborated in Sect. 4.5. Such a connection can be translated into efficient ROI defining rules. In [12,27] the ROI was defined based on the computed depths. Similarly, in [47] the upper part of the image is truncated. Here, however, the ROI's upper border line is adaptively determined, keeping only image rows that correspond to the desired distance range. In the special case of setting this distance bound to infinity, the computed horizon may also serve for upper image pruning [2].

Most road or lane estimation systems maintain some sort of tracking mechanism as discussed in Sect. 4.4. Knowing the detected location of the road or lane boundaries at previous frames and predicting the image movements in the present frame enables the definition of ROIs at image regions where these boundaries are expected to be found [6,35]. In [6] this approach is taken one step further: the lane boundaries are iteratively estimated using gradually increasing part of the image—from bottom to top. The ROIs for the first image segment (the lowest one) are determined using the tracking mechanism. Then, after each iteration, the most updated estimate of boundaries' location is utilized for defining an updated ROI for the next image part to be processed. It should be noted that beside assuming that the tracking algorithm is capable of modeling the motion between consecutive frames, these approaches also assume that the lastly detected boundaries are not erroneous. In case of a momentarily failure, recovery becomes rather challenging.

4.2 Bottom-up feature extraction

Once relevant image parts are determined, various features can be extracted. These features should contain the required information for the road or lane model fitting procedure that will follow. Throughout this paper the two tasks—road and lane estimation—are presented under a unified framework. Indeed, for most of the functional modules the distinction between the two tasks may be considered to be negligible. Here on the other hand, in the feature extraction module, totaly different features are sought for each of the estimation tasks. This difference stems from the different physical instruments being used for marking the boundaries of roads and lanes. Beside the confronted task, the choice of features tightly depends on the set of assumptions that we are willing to take. This section elaborates the features that were used for each of the estimation tasks and some observations are drawn regarding their underlying assumptions.

Lane boundaries features In general, lane boundaries are marked by different types of lane marks. In [3] the appearance variety of these lane marks is elaborated. Their shape may alter, from continuous lines through dashed lines and even circular reflectors may be encountered on some roads. Lane marks' color is also subject to change, e.g. white, yellow, orange, and cyan colors are all in use at different locations. Beside having various appearances, another challenge stems from the perspective effect which causes difficulties when trying to detect such narrow objects at distance.

Lane marks can be detected either based on their shape or color. The least restrictive assumption about the lane marks is that they have different appearance compared with the road. Such assumption leads to a whole family of gradient-based features and their variants. Simple gradients were computed in [7,50,51], either from the original image or from a smoothed version of it. Steerable filters were used by [3] which enable measuring the directional response at any direction by convolution with only three kernels. This property was utilized for the extraction of edges at presumed directions. According to feedbacks from the tracking mechanism the rough location and direction of the lane boundaries was predicted and high steerable filters responses were sought at these regions and directions. This type of filters is also capable of supplying the maximal and minimal responses with respect to examined directions. This property can be utilized for detecting circular reflectors: due to their characteristic isotropic response they tend to produce very small gap between the minimal and maximal responses.

In [2,4,6] the narrow shape of lane marks together with the assumption of having brighter intensity than their surrounding motivated the authors to search for low-high-low intensity pattern along image rows. Box filter was applied to the image in [2,6], while [4] convolved the image with a step filter and then searched for couples of adjacent responses with opposite signs. The raw responses of the aforementioned filters either serves directly as the extracted features (e.g. [50]) or, in some cases (e.g. [4,51]), go through a thresholding step that yields a binary edge map. In [51] the threshold was chosen adaptively according to the local average brightness in the image. In [11,20] a binary map was produced through a different technique. The reliability of the features was improved by dividing the image into 8×8 pixel blocks and classifying each of them as lane marks or not. Class determination was based on the conjunction of three criterions: first, the gap between the maximal and minimal intensity value should exceed some predefined threshold. Second, the distance between the two largest histogram bins was examined. These two criterions verify that the block contains both road and lane mark pixels. Last, the expected elongate shape of the lane marks is enforced by rotating the block in various angles and examine the ratio $var(x)/var(y)$. At first glance, this technique appears different from its box filters counterparts. However, it should be noted that the same set of assumptions underlies all these techniques, namely, brightness change and narrow shape of lane marks.

Whether we use gradient filters or box filters, the scale of the kernel should be determined. Due to the perspective distortion no single scale is suitable for the entire image. In some works (e.g. [2,21]) the scale of the kernel was adjusted for each row of the image according to the expected lane mark width, supported by the predicted depth at that row. A more commonly practiced technique [3,9,10,19,20,28, 51] circumvents the need for varying kernels by first warping the image in a manner that compensates the perspective effect. In this inverse-perspective image, sometimes referred to as "bird's-eye view", the lane marks width is equal at all distances. At the same time, the width of the lane and road becomes constant as well, a very advantageous property that facilitates model fitting (see Sect. 4.3). Moreover, in case that several sensors are available, such an orthographic image may serve as a convenient common ground for the fusion of their input images [20]. Inverse-perspective warping requires the system to be aware of the geometrical connection between the 2D image and the 3D ground plane. Methods to obtain and maintain this connection along the drive are elaborated in Sect. 4.5. In addition, this transformation is associated with some computational cost and a slight loss of resolution.

Another approach for lane marks' detection assumes that their brightness or color is known. LIDAR reflectance measurements were thresholded in [35]. In [10] the last three images were averaged and adaptive thresholding mechanism was applied. The averaging step has merits since dashed lane marks appear more continuous in the obtained image. In [5,21] the color distribution of lane marks was learned offline. Next, image pixels were classified according to the likelihood of belonging to a lane mark. The resulting features of all aforementioned procedures are sets of segmented blobs (with high enough intensity or color probability). It is very likely, however, that some portion of the image possesses similar colors even though it is not a lane mark. Hence, the set of extracted blobs is prone to contain some clutter and these algorithms are usually augmented with an additional filtering step. In [5], for example, these blobs were filtered primarily according to their size and shape. Then, for those blobs that could not be obviously classified as lane marks or as clutter, spatial location was tracked along few frames and the detected motion was compared with the ground plane motion, thus verifying that this blob lays on the ground. In [21] the response to a box filter was measured and utilized for improving the color-based segmentation. Here, however, instead of using this filter as a second step for clutter rejection the response image was fused with the color likelihood image into a unified score image that served as the final output of the feature extraction module.

Road boundaries' features Unlike lanes, roads are not always bounded by man-made markings. Depending on the type of road and environment, different cues will suggest the location of the road boundaries. Few examples may include curbs, usually found at urban environments, barriers which bound highway roads, and even dirt roads which have no marking at all and only color or textural change between road and off-road areas can indicate the boundaries. Due to this cues variety, no single feature is appropriate for all scenarios and different features were chosen for different systems according to the expected environments they should handle or according to the set of assumptions they may take.

In [2,8,32,36,37,39,46] it was assumed that there exists an elevation gap between the road and its surrounding. Such a cue leads to features that examine the three-dimensional structure of the observed scene. In [8,39] stereoscopic vision tools were used for the extraction of the scene structure. Next, surfaces with similar normal directions were clustered and curbs were detected. In [2,32,36,37] 3D data were directly collected by integrating a LIDAR sensor in the system. Here, the surface "roughness" (i.e. the elevation variance) is usually computed and serves as feature for segmenting the road. In [37], as an example, up to 5mm tolerance was defined for roughness-based segmentation mechanism. In [2] virtual rays were sent from the vehicle in various directions and the first roughness increase along each ray was detected and marked as road boundary. Such a method makes use of another assumption (very commonly taken) about the vehicle being currently located on the road.

While using the same assumption regarding the elevation differences, the authors of [46] have suggested an elegant method which circumvents the need for dense reconstruction of 3D scene points. Instead, estimates of the road plane position and the vehicle motion are maintained. This allows them to compute a homographic projection that warps the current frame to its preceding frame. However, since this homography was computed according to the road plane, only pixels residing on that plane are expected to match when comparing the previous image with the warped version of the current one. By examining the sum of absolute deviations between the two images road segmentation was accomplished.

A different approach for extracting road features is based on appearance rather than 3D structure. In this type of features it is assumed that the road has uniform appearance which is different from its surrounding. A region-growing algorithm was implemented in [49]. The colored image was first converted into an illumination-invariant intensity image. Then, seven seed points were placed at the bottom of the image, where road is assumed to be observed, and regions were grown. Similar assumption regarding the image bottom was taken in [3]. Here, a window was defined in the bottom of the image, where road presence is assumed. This "safe window" was then cropped and served as a template that was matched to further distal road parts.

Unlike [49] where intensity homogeneity was required and unlike [3] where the assumption about appearance constancy included the spatial distribution of the colors, in [29,30] the spatial aspect was dropped and only the color distribution of the road was utilized. Histograms were chosen in [30] for the distribution representation. Image colors were first transformed into several illumination-invariant color channels. Then, a safe window was defined at the image bottom and its pixels were accumulated in histograms according to their special channels intensities. Finally, the rest of the image was segmented according to the computed intensities' likelihood and a majority voting between the channels' segmentation output. A Gaussian Mixture Model was adopted in [29] for the description of road color distribution. Based on a safe window the parameters of this model were estimated. Next, color likelihoods were computed for image pixels and integrated in a Bayesian network framework. In this network every pixel was represented by a node and edges were defined according to pixels' adjacency. Hidden variables were also added to the network's nodes to account for possible shadow effects in the corresponding pixel. The Bayesian network framework enabled the integration and correct balance between these and additional factors such as boundaries' continuity and correct segmentation structure—i.e. it was enforced that each row is segmented into three parts: off-road, on-road, and then off-road again.

Textural features were examined by [25,26,28]. While the exact texture varies according to the material of the road, the footprints made by wheels of preceding vehicles add a strong directional component to the texture. Due to perspective effect the textural directions are not expected to be constant across the image of the road; nevertheless, under a straight road assumption they will all point to a common vanishing point. In [25,26] Gabor filters were utilized for the detection of the texture's dominant direction on each image point. Then, vanishing point was estimated through a voting scheme. Finally, the road was segmented by searching for the two most extrinsic rays that pass though the vanishing point and have high enough directional support. Similar scheme was implemented in [28] with different choice of textural features. Several alternatives were examined and it was found that Walsh–Hadamard features reduce computational overhead dramatically while keeping system accuracy almost unharmed.

The appearance approach for road segmentation can be strengthened: not only that the road has uniform appearance which is different from its surrounding, but also this appearance is known a-priori. This leads to classification-based algorithms such as those found in [5,12,28]. Preceding the aforementioned vanishing-point voting mechanism of [28], the extracted textural features were fed into an Adaboost classifier that was trained at preprocessing stage for road surface detection. In [5] typical colors of roads were learned a-priori and served as loose rules for road segmentation. Similar to [3] template-matching algorithm was also implemented in [12]. Here, however, the template was not cropped from the bottom of the image, but rather a set of predefined road templates was used while seeking the image for matching parts.

4.3 Road and lane model fitting

The road and lane detection is usually guided in a top-down manner by fitting a geometric model to the visual features extracted in a particular frame. Similar model fitting methods are used for both roads and lanes which are modeled as a 2D path with left and right boundaries either in the original headway view or in a virtual bird's-eye view created using the inverse perspective transformation.

The main goal of this stage is to extract a compact high-level representation of the path, which can be used for decision making. In the process, the noisy bottom-up path detection is improved by assuming a smooth path model with constraints on its width and curvature. This path representation is often further refined by matching to previous frames in the temporal integration stage (see Sect. 4.4). A path is usually represented by either its boundary points or by its centerline and lateral extent at each centerline location, which uniquely defines the boundaries. Transforming the frame to bird's-eye view simplifies the geometric model since the boundaries of the path become similar in curvature and the path's width is roughly constant. The models can be divided into parametric (e.g. lines [10,12,25,26,28,47]), semi-parametric (e.g. splines and poly-lines [2,50,52,53]) or non-parametric (e.g. continuous, but not necessarily smooth boundaries [27,29]). In most cases model parameter fitting has to cope with noisy boundary points extracted from the image in the form of missing data and a large relative amount of outliers. Random Sampling Consensus (RANSAC) [54] is commonly used for model fitting for all model types [2,7,10,20,51,53] thanks to its ability to detect outliers and fit a model to the inliers only.

For most methods the input data to the geometric model fitting module are a set of points extracted from the boundaries of the lane or the road. More information, however, may be computed in a bottom-up manner to guide the model fitting. In [53] points are grouped into lines. In [2] additional boundary direction information is added to each boundary point by computing a hessian filter. In [2,8,9] the distance transform is applied to the extracted boundary to obtain a smooth scoring for each pixel representing its distance from the closest boundary point. With such a representation it is straightforward to efficiently obtain a fitting score for a hypothesized curve model. Finally, in [2] the centerline model is fitted to a combined probability map computed from LIDAR and camera evidences.

Parametric models The simplest geometric models used for the path boundaries are straight lines [10,12,25,26,28,47] which are a good approximation for the short range and as the most common case in highway scenarios. Curved roads were modeled in a bird's eye-view using parabolic curves in [2,3] and using generic circumference arcs in [19,51]. To handle the more general curved paths observed in the projective headway view [6,7] use parabolic curves and [4] use hyperbolic polynomial curves, although semi-parametric models such as splines are recently more commonly used for this purpose.

Various methods are used to fit the parametric model to the path boundary evidence. Both [3,19] use least squares optimization. References [2,10,51] use RANSAC together with a least squares optimization to improve robustness to outliers. In [2] this method is used to fit to centerline evidence and an additional step is taken: the final inlier set is divided into connected components, where a point is connected to all points in a 1-m ball around it in ground distance and the largest component is taken as the inlier set to which parabolic model is fitted. This prevents parabola fitting across multiple centerlines, by requiring that an entire identified centerline is connected. Labayrade et al. [4] use the weighted least squares algorithm [55] derived from the M-estimators theory to obtain robust curve estimations. In [7] the curvature and orientation of the boundaries are found using a generalized hough transform while the position of each boundary is fitted to the image using a genetic algorithm.

Additional fitting methods are specialized to linear model matching. The Hough transform [56] is used in [10,19]. In the direct view the vanishing point can be used as an anchor to both linear boundaries assuming a constant path width. First the vanishing point is detected using either the Hough transform [47], or by direct voting of oriented edglets [25,26]. Among the lines passing through the vanishing point, the two with the best support are chosen as the boundaries. In the bird's-eye view, further assuming that the car is parallel to the road (and therefore that the lines are parallel to the y-axis) a simple integration over the y axis can be used to detect the position of the lanes in the x-axis [28]. To loosen the assumption regarding car and lane parallelism, Pomerleau [12] tries several rotations of the bird's-eye view and uses the sharpness of the y-integration result to also detect the angle of the road ahead.

Semi-parametric models The advantage of semi-parametric models is that they do not assume a specific global geometry of the path. On the down side fitting the model must be carefully carried out to prevent over-fitting and unrealistic path curvature. In [5,11] the headway image is split to horizontal strips and a constrained Hough transform is used to find the best linear fit to each lane mark in each slice, resulting in a piece-wise linear model. In [11] a similar model is fitted over time in bird's-eye view by fitting a linear model to the short range at each time step.

Splines are smooth piecewise polynomial functions, and they are widely used in representing curves. Different spline models with different properties were used to model the lane boundaries/centerline. Kim [53] use Cubic Splines because the curve includes the control points, Wang et al. [52] use B-Splines and [50] use Active Contours (Snakes) since they enable energy-based optimization. Huang et al. [2] use a Cubic Hermit Spline which ensures that the extracted tangents at lane points are continuous between pairs of control points. In all spline models the curve is parameterized by a set of control points either on (e.g. [53]) or near (e.g. [52]) the curve. The advantage of splines versus parametric models (e.g. parabolas) is that a small change in the parameters is correlated to a small change in the curve, facilitating the use of tracked control points from the previous frame for model initialization in the current frame.

In lane detection there are usually many lane marking detections that can be control point candidates, and choosing among them is a delicate issue. The number of points affects the curve complexity; they should be evenly distributed along the curve to prevent high curvatures, more confident features should be preferred and finally the fitted spline should have good support from the remaining features. In [53] a RANSAC approach is used to generate a set of 100 hypothesis splines: first the lane features are grouped into lines and then 2, 3 or 4 control points are heuristically chosen from 1, 2 or 3 line ends. The points are chosen to be as equally spaced as possible. The number of control points are randomly selected per hypothesis. The hypotheses are evaluated based on their lane-marking support plus a penalty score for high curvature changes at points with no lane-marking evidence between them. In [2] the algorithm first selects 100 "seed" features at close range (in bird-eye view) since they are usually less noisy. A "greedy" search is then used to find each one of the next points by evaluating the resulting spline. At each step only points distanced around 50 pixels from the previous point are considered to generate evenly spaced control points.

Non-parametric models Less common are non-parametric line models, demanding only that the line is continuous but not necessarily differentiable. In [27] an Ant Colony Optimization (ACO) is used for finding optimal trajectories on the image plane, starting at the lower sides of the image and ending at the vanishing point. Trajectory scoring in this case is according to the alignment with the image edge-map. In [29] a learned Hierarchical Bayesian network model is used. Each image row is segmented into three segments of road/non-road/road using two thresholds, in a way that softly enforces continuity between consecutive rows. In [9] an original lane model is used which matches individual left and right boundary points in the bird's-eye view. The model is

fitted using a particle filter approach, progressing along the Y axis from the bottom of the image upwards (that is, the Y axis is treated as a time axis).

Lateral model Different assumptions are made by existing methods on the lateral extend of the road/lane. The strongest assumption is that the lane width is known (between 3–3.5 m). Notice that under this strong assumption the left- and right-lane lines are fully coupled, so the number of parameters to be estimated for a lane is cut by half compared with independent estimation of the lines. In [2] a known lane width is assumed and used in a probabilistic centerline detection framework. A slightly weaker assumption is that the real-world width of the road/lane is roughly constant (but not necessarily 3 m or any other known quantity). This assumption is true in most scenarios, and yet it is strong enough to cut the number of parameters in approximately half (one more parameter of the lane width has to be estimated). Left- and right-lane detection are constrained in such a manner in [4,6]. Methods using the vanishing point as an anchor [25,26,47] assume constant lane width too. In [3,12,20] a road/lane section model is used for guiding boundary detection, which also implicitly assumes constant road and lane width.

Notice, however, that even when detecting a single road/lane, the imaged width can slightly be increasing or decreasing due to the pitch movement of the vehicle, presence of up/down-hills and/or changing lane/road width. To compensate for slight changes in lane width [53] incorporate a linear width change assumption (in a single image) within their Bayesian model. In [9] an even weaker assumption of continuous lane width is used.

Model complexity As always when fitting a model to noisy and partial data there is a tradeoff between over-constrained models which do not cover all existing geometries and under-constrained ones which tend to over-fit noisy features. Kim [53] generate splines with different numbers of control points, corresponding to models with different complexity levels, which are scored both by rewarding lane marking support and by penalizing for curve direction changes without lane marking support. In [19] digital map+GPS is used to distinguish curved versus linear road regions, applying the more complex curved model only when needed.

In general, the lane and road modeling problem is characterized by a 'long tail' of exceptions, which cannot be neglected due to the high reliability demands. In such circumstances, choosing the model family a priori is sub-optimal, and on-line model selection can have considerable advantages. On-line model selection should try different models and score them by weighting the model complexity with its fitting score. Formal approaches in which such considerations can be done are Bayesian model inference [57], Minimum Description Length (MDL) [58], or Structural risk minimization (SRM) [59].

Beyond the single lane/road model The vast majority of the methods assume line lane/road topology which does not include merging, splitting and ending lanes or roads. The exceptions are [53] (lanes) and [2] (roads,lanes) which are designed to deal with non-linear road and lane topologies. In [19] lanes adjacent to the host lane are detected by extrapolating to the side of the host lane with its detected width. In [18] a solution is proposed to the lane assignment problem: given the number of lanes (from a digital map source+GPS) identify in which lane the vehicle is.

4.4 Temporal integration

Integrating knowledge from previous frames has three goals: improving the accuracy of a correct detection, reducing the required computation and correcting erroneous detections. Detection accuracy is improved by predicting the detection and smoothing the result over time. Computation is reduced by supplying a good initialization of the model parameters and constraining the searched parameter space, as well as limiting the image regions from which features are extracted [4]. Correcting erroneous detections can be achieved by comparing the current detection with previous ones and rejecting the less likely one in case of large discrepancy [2].

Temporal integration can be done by 2D lane/road model tracking in headway view. For example, in [50] an inertial energy propagated from the previous frame serves as a tracking mechanism for the 2D active contours (snakes) lane model. However, the more common solution is to track the lane/road model in real-world coordinates. The lane model is transformed from image to real-world coordinates, typically using an inverse perspective transformation. Estimation of the vehicle dynamics in the real world is used to predict the position of the previously detected lane in the current frame. Such vehicle motion estimation can be obtained in several ways: using visual input to match consecutive frames and compute an ego-motion model [46], using car odometry (speed and yaw rate) [9] and most accurately by combining GPS and IMU information [2,22,23]. The transformed expected lane is then combined with lane evidence from the current frame to find the best estimate for current lane/road model parameters. Most methods use either Kalman Filter [4,3,10,53] or Particle Filtering [8,9,22,53] for this task.

A weak point in the mostly used tracking methods is the naive inverse perspective transformation, assuming stable camera calibration and flat zero-level ground surface. Vehicle vibrations and changing ground slopes inject noise and abrupt changes into the time series, posing some difficulty for tracking methods with strong smoothness assumptions (e.g. Kalman Filter).

In [53] a Gaussian model for the car motion is used to sample 50 motions, each predicts a position of the lane marks in the current frame, and used in a particle filtering approach. Tracking a lane mark described by a spline as in [53] poses the following problem: each of the control points is moved by the motion model, some ending up behind the vehicle and outside the image plane, requiring rearrangement of the control points. In [9] each lane is described by equally spaced (left–right) pairs of lane boundary points which are tracked in a single frame along the lane using particle filtering. In the next frame these points are tracked using car odometry and the particle filtering process is used to efficiently find only the pairs of points on the newly revealed part of the lane.

Other models are also used for tracking multiple hypotheses, tracking dense representations of the lanes or tracking low-level features. In [53] lateral and temporal coherence jointly guide a probabilistic model for grouping the two lane boundaries, in which the temporal reasoning is based on Dynamic Bayesian Networks (DBN) [60] and used to score hypothesis generated in the previous frame. In [2] a dense centerline probability map is updated from the previous frame, and knowledge learned on road regions is propagated to the current frame. In [3,12] a road section template is adapted online in each new frame. In [46] the found road region is projected from the previous frame to the current one using estimated ego-motion to guide road region detection in the current frame. Katramados et al. [30] uses temporal filtering for rejecting non-stable segments of clear road path.

4.5 Image to world correspondence

Knowing the geometrical connection between the two-dimensional image and the three-dimensional environment requires the estimation of the camera position and orientation with respect to the ground surface. Such a piece of information may come in useful at all stages of the road or lane estimation task. At the image pre-processing stage, for instance, it enables rejection of the image part above the horizon [2]. Producing an inverse perspective image is also enabled due to the understanding of camera and world connection. Here, the homography which connects the image plane with the ground plane is estimated and defines the warping transformation. Such a perspective-free view is very beneficial both for feature extraction and for model fitting as described in Sects. 4.2 and 4.3, respectively. Finally, the image to world correspondence enables tracking of the vehicle state in the 3D world rather than just tracking the road/lane appearance in the image, thus allowing integration of other real-world measurements coming from different sources.

Different techniques were used to obtain the correspondence between video frames and the 3D world. In the most restrictive form, it was assumed by [2] that this connection is kept constant along the drive and thus it was calibrated beforehand. This strong assumption, however, was found by the authors as problematic in scenarios where the ground slope changes suddenly.

In [4,51] only the pitch of the camera with respect to the ground plane was estimated. In [51] the road's vanishing point was detected and its image height was used for the computation of the sought pitch angle. In [4], the pitch angle was concatenated to the 3D model parameters vector. Then, the connection between this vector and the 2D lane boundaries' location and direction was learnt. As a result, the pitch angle was estimated together with the rest of the model parameters, given the 2D measurement of the examined frame.

Usually, the inverse perspective image is only an outcome of this image to world connection. Nevertheless, in [28] both camera pitch and yaw angles were estimated using this special view. Each hypothesis for this couple of angles leads to a different warped image and a couple that produces bird's-eye view with two distinctive peaks was sought.

Structure from motion techniques are also a possibility for obtaining the camera pose with respect to the ground plane. In [46] Harris corners were tracked across the video frames, from which the essential matrix was computed using a robust variant of the eight points algorithm. This essential matrix encapsulates the camera ego-motion. Then, triangulation method was applied to the tracked features and a plane was robustly fitted to their 3D reconstruction. It should be noted that such a technique assume the presence of enough distinctive and easy to track feature points on the road surface.

Most of the aforementioned techniques make use of a planar world assumption. In [40] this assumption was dropped and the 3D profile of the road was reconstructed from a disparity image produced from a stereo vision system.

5 System level integration

As seen in the previous section, there are several algorithmic methods for each processing step in the road and lane detection system. Often, different methods rely on different assumptions and sometimes on different modalities. Some works use multiple algorithms and modalities in the same processing step to achieve enhanced robustness [2,4,19,21–24,28], and in general such fusion of multiple information sources seems to be critical to achieve reasonable system level performance.

Different algorithms for the same task can be combined by running all of them in parallel and weighing their results (model averaging), or by choosing which one to apply in given circumstances (model selection). From a decision theoretic point of view, running all the algorithms and weighing their response is the best policy (as long as each of the algorithms has non-zero probability to be correct in its assumptions) [61]. In [4] two different algorithms for lane fitting are

employed, and only if they agree (within some tolerance) the lane found is used. Otherwise, low confidence is reported. In [2] two different lane fitting algorithms, as well as a third LIDAR based algorithm are all combined in a single probability map. Different modalities, mostly LIDAR with global positioning (GPS+IMU+map) were successfully combined in many urban challenge papers [2,21–24,32]. However, running several algorithms is typically computationally costly.

An alternative to running all algorithms is to apply them serially, that is, use the second when the first fails and reports low confidence. This is most reasonable when the next algorithm uses weaker assumptions than the previous one which failed. For example, in [28] an algorithm for road classification assuming a priori known road characteristics is used as a default. When it reports low confidence, the system switches to a second road segmentation algorithm, which only assumes that the road region is different from its surroundings. In [19] lane model selection is done based on GPS and map information: the global positioning information dictates the choice between fitting a straight or curved lane model.

Another issue of critical importance for robust system is the reporting of confidence intervals, both by its algorithmic constituents and by the system as a whole. This is true both for open-loop warning systems and for closed-loop autonomous systems. With warning systems, a warning is issued when an event is detected with high confidence. However, when confidence is low the system will typically only report its low confidence and take no further action [4,28]. The key point here is "do no harm": while obviously undesirable, low confidence can be tolerated if identified correctly and the driver is made aware of that.

In closed-loop autonomous driving systems, the system has greater responsibility and low confidence is harder to tolerate. Still, even in such system, identified low confidence is better than an error. Typically, confidence is computed based on some model fitting score, and low confidence of an algorithm indicates that its assumptions do not hold currently. Hence when low confidence is reported, the system can switch to another algorithm [28] or give smaller weight to its voting if it runs in parallel to other algorithms. In a similar fashion, when single frame road or lane detection fails and reports low confidence, the system can use road and lane models tracked from earlier time frames. In [2] the system reported low confidence lane information 1 m in front of the host vehicle for 35% of its driving time. Nevertheless, it was able to keep staying on track by tracking high confidence measurements from previous frames.

6 Evaluation

The ability to benchmark and evaluate algorithms is necessary to compare performance of different techniques, asses system's maturity and identify their weak spots. However,

this issue is highly problematic in the lane and road detection literature, due to the lack of accepted test protocols, performance metrics and public data sets. Many papers report their results only qualitatively [4,5,7,9,11,12,46,50,53]. Some of these papers report hours of successful driving [4,7,12], but without ground truth information for these experiments no quantification of the result is possible. In addition, almost no two papers use exactly the same metric, and comparison is nearly impossible [10] (see [28] for an exception).

Many papers focus on a specific sub-task of the lane/road detection system and measure their success with metrics specific for the relevant aspect. In [26], for example, a technique for the estimation of vanishing point is suggested and performance is measured using the mean square error of the vanishing point estimation. In [28] the same task is judged by the angle between the lines connecting the host vehicle to the real and estimated vanishing points. A mean angle of 1.7 degrees is reported. In [18] a lane assignment problem is discussed, where the task is to classify the host lane into one of four existing lanes. Performance is measured using a confusion matrix, with error rates lower than 10% at all conditions.

Papers handling road segmentation sometimes judge their success by counting per-pixel binary classification (road vs. non-road) errors. Per-pixel ROC curves are drawn in [27,28], and the percentage of correct pixels (97.8%) is reported in [30]. The work presented in [62] suggests a method for fitting a parametric lane model, and its performance is judged by the mean square error of the parameters estimated. The interesting aspect in this work is in the way ground truth of the parametric model was achieved: the experiments were carried using computer graphics simulated road images.

While the aforementioned papers present evaluation of specific subtasks, it is very hard to relate the reported performance to the performance of a whole system. In contrast, there are papers presenting a full system with quantitative performance evaluation [2,3,6,10,28]. In this group, some papers [2,3] report mainly detection accuracy statistics, while others report probability of successful detection, where success is defined by a fixed threshold on the estimation accuracy [6,10,28]. While the latter approach is more complicated (success probability is only defined for a given accuracy, and hence it depends on some hidden parameters), it is the most relevant to estimation of the system reliability. The reason is that when accuracy reaches a certain level (the level allowing correct system performance), its statistics are not important anymore. The important quantity is the percentage of cases in which this accuracy is not achieved.

Specifically, in [2] statistics of lane centerline estimation are reported. Mean accuracy of $57-70$ cm is reported for distance 1–50 m from the vehicle, respectively, in complicated urban conditions. In [3] much better results are reported for highway roads (mean error of 8.2 cm with standard deviation of 13.1 cm). However, the results are not directly compara-

ble due to differences in the data set used and the details of error estimation. The latter paper [3] contains also a good discussion of possible performance measurements, including the error in estimation of lateral velocity, which is mostly relevant to the LDW feature.

An impressive system-level evaluation of road detection is presented in [28]. Successful road boundary detection is defined as having accuracy better than 30 cm for the boundary 6 m ahead of the host vehicle (which is actually measured in image coordinates as distance of less then 18 pixels at a certain image row). The system is reporting availability (i.e. having high confidence) at 92–100% of the time, with error probability of 0.005–0.02 in various conditions (error probability is measured only in the high confidence intervals). In [6,10] similar scores are reported for the task of lane detection. A success probability of 0.99 is reported in [6] for accuracy in distance 15 m from the vehicle. Interestingly, the reported night performance (0.994) is higher than day performance −0.988 (similar phenomena is reported in [3] when night and noon performance are compared). In [10] success probabilities of 98–99% are reported for lane detection on highways, dropping to 86% in city driving.

Part of the difficulty in quantitative evaluation is in obtaining ground truth for the data. Human annotation is used for small-medium data sets [10,26]. Alternatively, additional equipment on the test vehicle can be used to collect ground truth. In [2] map data and global positioning information are used to provide rough ground truth, which is refined by human corrections. The test vehicle used in [3] uses specific cameras looking down on the vehicle's sides to collect ground truth lane information.

DARPA's 2005 Grand Challenge [25,27–29] and 2007 Urban Challenge [2,20–24,32] were major test-beds for limited autonomous driving systems. However, this proof of concept is very limited due to reasons mentioned in Sect. 2: participants of the 2005 challenge did not have to cope with lane issues at all, and most of the groups in the 2007 challenge used very limited on-board perception and relied on exact map information instead. With some exceptions [2,28], most of the papers describing the participating vehicles do not provide quantitative system-level performance data for on-board perception.

As can be seen from this non-exhaustive summary, the variety of metrics and data sets used makes it very hard to draw conclusion regarding algorithm quality or whole system's maturity level. Two encouraging exceptions are the papers [2,10], who published their data on the web. An accepted public benchmark with well-defined evaluation protocols can have a large positive impact on current circumstances. Such a benchmark should contain a large set of video streams with ground truth, preferably augmented with additional sensors such as LIDAR and GPS+GIS. Similar to the database presented in [63], such a database should

also be annotated with meta variables stating the scenario type, according to the cases discussed in Sect. 2. Finally, we point out that such an annotated database will also be useful for machine learning training, thus contributing to algorithm development.

7 Concluding remarks

Considerable progress in road understanding has been made in the past years, with two powerful engines pushing forward: Lane Departure Warning (LDW) systems which are turning into commercial products, and the DARPA challenges for fully autonomous driving. This focused the research attention on the two extreme poles of the road and lane understanding problem: the most simple problem of LDW (single lane, short distance) and the hardest problems (fully autonomous driving in desert and urban environments). This in turn led to the development of several very different approaches to road understanding. LDW systems were developed into complex vision-based systems with some high-level reasoning, allowing some reliability to different conditions. The fully autonomous dominant solution (with some exceptions [2,20]) was to give up on-board full road perception, and rely instead on integration of very accurate global positioning information, obtained from GPS and IMU, with high-resolution maps and aerial images. On board perception in these systems is usually limited to LIDAR-based verification of the fine positioning information. For the cross-country autonomous driving the solutions focused on demarcation of very near road structure, often 10–15 m in front of the vehicle using simple and robust road models.

This research distribution has left a large gap in middle-complexity road understanding issues: perception of multiple lanes and non-linear topologies of lane and road. The research conducted towards full autonomy does not give answers to these middle complexity problems. Instead, since the full autonomous problems are very hard, ad-hoc solutions are developed which bypass the need for comprehensive onboard perception (in the Urban Challenge case), or focus on very limited problem aspects (near road segmentation for the desert challenges). However, the less-researched middle complexity lane and road understanding are the most required capabilities for further advance in commercial active safety features.

The challenges for research in the near decade are mainly of two types: extend the scope of road understanding and increase its reliability. The first challenge is to extend current road and lane detection abilities into new domains. As can be seen from Table 1, features of the next generation require understanding of multiple lanes, far ahead and in the back of the host vehicle, as well as identification of lane and road splits and merges. Some examples of these challenging conditions are shown in Fig. 3. This challenge requires
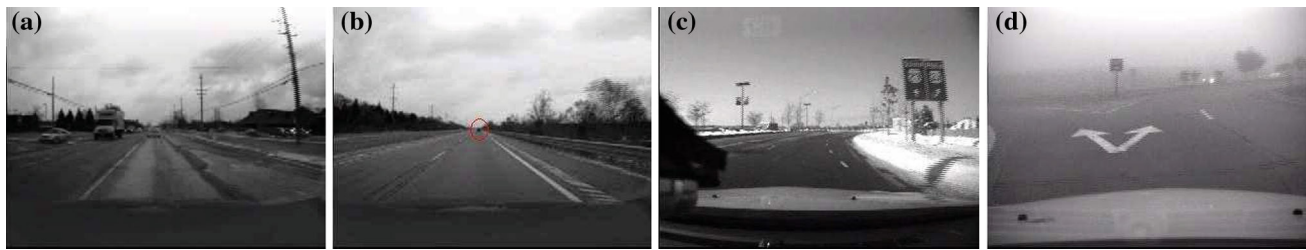
Fig. 3 Road and lane challenges for near-future features: **a** Identifying multiple lanes and their traffic directions, required for lane change assist. **b** Identifying lanes in the distance is required to determine if a far vehicle is in the host lane. This information is needed for overtake planning. **c** Detection of lane splits is required for lane centering. **d** Road split detection is needed for autonomous navigation

the development of new road scene representations, which are rich enough to describe multiple lanes with non-linear topology, and yet can be reliably extracted and tracked from a video stream.

The reliability challenge is harder than the first, at least for systems based primarily on vision. The reliability of current systems, which is enough for warning systems, may not be enough for closed-loop features, requiring error rates which are often orders of magnitude lower. While relatively simple algorithms with simple assumptions may work in a large majority of the cases, a complete system with high reliability must include a mixture of algorithms relying on different assumptions and information sources. The algorithms should operate in parallel and get weighed, or alternatively be applied serially, with complex algorithms activated only when assumptions of simpler ones failed. Moreover, the system should be able to explicitly infer failure cases of algorithms' assumptions, in order to weight them appropriately (when all algorithms are run), or switch between them in the serial case.

Due to the high-reliability demands, building a vision-based system, even for the simplest applications, is a big development effort. Many functional blocks are required for a stable system, and many different conditions and assumptions have to be identified and handled. In addition a large validation effort is required, as many of the failure cases are rare and hard to predict. The high system complexity and the large development effort required to build reliable systems significantly constrain research and development in road understanding. What can we do to enable better progress under these conditions? Several lines of action may be fruitful:

– **Use modalities other then vision when possible** It has been demonstrated several times that problems which are hard for a pure vision-based system are much easier with other modalities. Road segmentation is much easier with LIDAR. When accurate and updated area maps are available, global positioning with GPS and IMU provides enormous simplification for the on-board lane sensing. It might be possible to use publicly available data as Google satellite images or StreetView for this purpose.

– **Adopt machine learning techniques** In some machine vision tasks, as well as other applicative domains (e.g. mail filtering or speech recognition), machine learning (ML) techniques enabled achieving more accurate results with less engineering effort. Lane and road understanding may be a convenient domain for ML techniques as unsupervised data in large quantities can be gathered almost for free, and large quantities of supervised data can be gathered with some effort by driving with additional sensors (LIDAR, GPS, IMU and maps) [2,3]. Machine learning techniques are already used for road classification in some cases [5,28], but larger functional blocks can also be addressed. One example in this direction is presented in [29], in which road segmentation is cast as parameter learning in a hierarchical Bayesian network. The method described was used in the winning vehicle of the DARPA 2005 challenge. It may be fruitful to approach more complex problems of multiple lane detection using similar structured output ML techniques.

– **A public benchmark** A big challenge of current research is the inability to compare performance of different methods due to the lack of public annotated benchmarks. Putting forward a large public video benchmark may reduce evaluation costs and enable cross-publication comparisons. Similar data sets increased progress rate in other domains such as pedestrian detection [64] or object class recognition [65].

References

1. Thorpe, C., Hebert, M., Kanade, T., Shafer, S.: Toward autonomous driving: the CMU Navlab. Part I: perception. IEEE Expert **6**, 31–42 (1991)
2. Huang, A.S., Moore, D., Antone, M., Olson, E., Teller, S.: Finding multiple lanes in urban road networks with vision and LIDAR. Auton. Robots **26**, 103–122 (2009)
3. McCall, J., Trivedi, M.: Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. IEEE Trans. Intell. Transp. Syst. **7**, 20–37 (2006)
4. Labayrade, R., Douret, J., Laneurit, J., Chapuis, R.: A reliable and robust lane detection system based on the parallel use of three algorithms for driving safety assistance. IEICE Trans. Inf. Syst. E **89D**, 2092–2100 (2006)

5. Cheng, H., Jeng, B., Tseng, P., Fan, K.: Lane detection with moving vehicles in the traffic scenes. IEEE Trans. Intell. Transp. Syst. **7**, 571–582 (2006)

6. Wu, S., Chiang, H., Perng, J., Chen, C., Wu, B., Lee, T.: The heterogeneous systems integration design and implementation for lane keeping on a vehicle. IEEE Trans. Intell. Transp. Syst. **9**, 246–263 (2008)

7. Samadzadegan, F., Sarafraz, A., Tabibi, M.: Automatic lane detection in image sequences for vision based navigation purposes. In: ISPRS Image Engineering and Vision Metrology (2006)

8. Danescu, R., Nedevschi, S.: Probabilistic lane tracking in difficult road scenarios using stereovision. IEEE Trans. Intell. Transp. Syst. **10**, 272–282 (2009)

9. Jiang, R., Klette, R., Vaudrey, T., Wang, S.: New lane model and distance transform for lane detection and tracking. In: Computer Analysis of Images and Patterns, pp. 1044–1052 (2009)

10. Borkar, A., Hayes, M., Smith, M.: Robust lane detection and tracking with RANSAC and Kalman filter. In: International Conference on Image Processing, pp. 3261–3264 (2009)

11. Shi, X., Kong, B., Zheng, F.: A new lane detection method based on feature pattern. In: International Congress on Image and Signal Processing, pp. 1–5 (2009)

12. Pomerleau, D.: RALPH: Rapidly adapting lateral position handler. In: IEEE Intelligent Vehicles Symposium (1995)

13. Zhou, T., Xu, R., Hu, X.F., Ye, Q.T.: A lane departure warning system based on virtual lane boundary. J. Inf. Sci. Eng. **24**, 293–305 (2008)

14. Burzio, G., et al.: Investigating the impact of a lane departure warning system in real driving conditions: a subjectivefield operational test. In: European Conference on Human Centred Design for Intelligent Transport Systems (2010)

15. Barickman, F.S., Smith, L., Jones, R.: Lane departure warning system research and test development. In: NHTSA Paper number 07-0495

16. Batavia, P.H.: Driver-adaptive lane departure warning systems. CMU-RI-TR-99-25 (1999)

17. Hofmann, U., Rieder, A., Dickmanns, E.: Radar and vision data fusion for hybrid adaptive cruise control on highways. Mach. Vis. Appl. **14**(1), 42–49 (2003)

18. Gao, T., Aghajan, H.: Self lane assignment using egocentric smart mobile camera for intelligent GPS navigation. In: Workshop on Egocentric Vision, pp. 57–62 (2009)

19. Jiang, Y., Gao, F., Xu, G.: Computer vision-based multiple-lane detection on straight road and in a curve. In: Image Analysis and Signal Processing, pp. 114–117 (2010)

20. Lipski, C., Scholz, B., Berger, K., Linz, C., Stich, T., Magnor, M.: A fast and robust approach to lane marking detection and lane tracking. In: Southwest Symposium on Image Analysis and Interpretation, pp. 57–60 (2008)

21. Kornhauser, A.L., et al.: DARPA Urban Challenge Princeton University Technical Paper (2007). http://www.stanford.edu/~jmayer/papers/darpa07.pdf

22. Urmson C., et al.: Autonomous driving in urban environments: boss and the urban challenge. J. Field Robot. **25**(8), 425–466 (2008)

23. Bacha, A. et al.: Odin: team victor Tango entry in the DARPA urban challenge. J. Field Robot. **25**(8), 467–492 (2008)

24. Montemerlo, M. et al.: Junior: the Stanford entry in the urban challenge. J. Field Robot. **25**(8), 569–597 (2008)

25. Rasmussen, C., Korah, T.: On-vehicle and aerial texture analysis for vision-based desert road following. In: CVPR Workshop on Machine Vision for Intelligent Vehicles, vol. III, p. 66 (2005)

26. Kong, H., Audibert, J., Ponce, J.: Vanishing point detection for road detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 96–103 (2009)

27. Broggi, A., Cattani, S.: An agent based evolutionary approach to path detection for off-road vehicle guidance. Pattern Recognit. Lett. **27**, 1164–1173 (2006)

28. Alon, Y., Ferencz, A., Shashua, A.: Off-road path following using region classification and geometric projection constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. I, pp. 689–696 (2006)

29. Nefian, A., Bradski, G.: Detection of drivable corridors for off-road autonomous navigation. In: International Conference on Image Processing, pp. 3025–3028 (2006)

30. Katramados, I., Crumpler, S., Breckon, T.: Real-time traversable surface detection by colour space fusion and temporal analysis. In: Computer Vision Systems, pp. 265–274 (2009)

31. Borkar, A., Hayes, M., Smith, M., Pankanti, S.: A layered approach to robust lane detection at night. In: IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, pp. 51–57 (2009)

32. Kammel, S., Pitzer, B.: LIDAR-based lane marker detection and mapping. In: IEEE Intelligent Vehicles Symposium, pp. 1137–1142 (2008)

33. US Department of Transportation, Federal Highway Administration, O.o.I.M.: Highway statistics (2005)

34. von Reyher, A., Joos, A., Winner, H.: A LIDAR-based approach for near range lane detection. In: IEEE Intelligent Vehicles Symposium, pp. 147–152 (2005)

35. Ogawa, T., Takagi, K.: Lane recognition using on-vehicle LIDAR. In: IEEE Intelligent Vehicles Symposium, pp. 540–545 (2006)

36. Takagi, K., Morikawa, K., Ogawa, T., Saburi, M.: Road environment recognition using on-vehicle LIDAR. In: IEEE Intelligent Vehicles Symposium, pp. 120–125 (2006)

37. Hernandez, J., Marcotegui, B.: Filtering of artifacts and pavement segmentation from mobile LIDAR data. In: Laser09, p. 329 (2009)

38. Cremean, L.B., Murray, R.: Model-based estimation of off-highway road geometry using single-axis LADAR and inertial sensing. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1661–1666 (2006)

39. Pradeep, V., Medioni, G., Weiland, J.: Piecewise planar modeling for step detection using stereo vision. In: Computer Vision Applications for the Visually Impaired (2008)

40. Sach, L., Atsuta, K., Hamamoto, K., Kondo, S.: A robust road profile estimation method for low texture stereo images. In: International Conference on Image Processing, pp. 4273–4276 (2009)

41. Michael, W., Aaron, E., Loren, K.: Consumer-grade global positioning system (GPS) accuracy and reliability. J. Forestry **103**, 169–173 (2005)

42. Caron, F., Duflos, E., Pomorski, D., Vanheeghe, P.: GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. Information Fusion, pp. 221–230 (2004)

43. Yi, Y.: On improving the accuarcy and reliability of GPS/INS based direct sensor georeferencing. Ph.D. dissertation, Ohio State University, Columbus (2007)

44. Ma, B., Lakshmanan, S., Hero, A.O., I.: Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion. IEEE Trans. Intell. Transp. Syst. **01**, 135–147 (2000)

45. Kaliyaperumal, K., Lakshmanan, S., Kluge, K.: An algorithm for detecting roads and obstacles in radar images. IEEE Trans. Vehicular Technol. **50**, 170–182 (2001)

46. Yamaguchi, K., Watanabe, A., Naito, T., Ninomiya, Y.: Road region estimation using a sequence of monocular images. In: International Conference on Pattern Recognition, pp. 1–4 (2008)

47. Zhang, G., Zheng, N., Cui, C., Yan, Y., Yuan, Z.: An efficient road detection method in noisy urban environment. In: IEEE Intelligent Vehicles Symposium (2009)

48. Mobileye homepage. http://www.mobileye.com/manufacturer-products/brochures

49. Alvarez, J., Lopez, A., Baldrich, R.: Shadow resistant road segmentation from a mobile monocular system. In: Iberian Conference on Pattern Recognition and Image Analysis, **II**, 9–16 (2007)
50. Sawano, H., Okada, M.: A road extraction method by an active contour model with inertia and differential features. IEICE Trans. Inf. Syst. E **89D**, 2257–2267 (2006)
51. Nieto, M., Salgado, L., Jaureguizar, F., Arrospide, J.: Robust multiple lane road modeling based on perspective analysis. In: International Conference on Image Processing, pp. 2396–2399 (2008)
52. Wang, Y., Teoh, E., Shen, D.: Lane detection and tracking using b-snake. Image Vis. Comput. **22**, 269–280 (2004)
53. Kim, Z.: Robust lane detection and tracking in challenging scenarios. IEEE Trans. Intell. Transp. Syst. **9**, 16–26 (2008)
54. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
55. Tarel, J.P., Ieng, S.S., Charbonnier, P.: Using robust estimation algorithms for tracking explicit curves. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I, pp. 492–507. Springer, London (2002)
56. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**, 11–15 (1972)
57. Liang, F., Troung, Y., Wong, W.: Automatic Bayesian model averaging for linear regression and applications in Bayesian curve fitting. Stat. Sin. **11**, 1005–1029 (2001)
58. Rissanen, J.: An introduction to the MDL principle (2008). http://www.mdl-research.org/jorma.rissanen/pub/Intro.pdf
59. Kohler, M., Krzyzak, A., Schäfer, D.: Application of structural risk minimization to multivariate smoothing spline regression estimates. Bernoulli, **8**(4), 475–489 (2002)
60. Dean, T., Kanazawa, K.: A model for reasoning about persistence and causation. Comput. Intell. **5**, 142–150 (1989)
61. Claeskens, G., Hjort, N.L.: Model Selection and Model Averaging. Cambridge University Press, Cambridge (2008)
62. Lopez, A., Serrat, J., Canero, C., Lumbreras, F.: Robust lane lines detection and quantitative assessment. In: Iberian Conference on Pattern Recognition and Image Analysis, vol. I, pp. 274–281 (2007)
63. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: a benchmark. In: CVPR (2009)
64. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, pp. 886–893. ACM, New York (2005)
65. Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 594–611 (2006)

Author Biographies
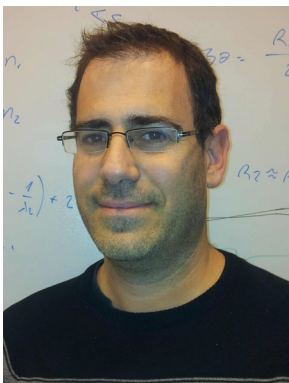
Aharon Bar Hillel received his B.A in mathematics and philosophy from Tel Aviv University and his Ph.D. in neuroscience from The Hebrew University of Jerusalem. He had worked for Intel Research from 2006–2008 and since 2008 he is a researcher in the Smart Sensing and Vision Systems Group at General Motors, Advanced Technical Center-Israel. He is a machine learning and computer vision practitioner.

Ronen Lerner received his B.Sc degree in computer science and mathematics from the University of Haifa, and his M.Sc and Ph.D. degrees from the Technion, Israel Institute of Technology. Currently he is a senior researcher in the Smart Sensing and Vision Systems Group at General Motors, Advanced Technical Center-Israel. His research interests are in image-processing and computer-vision.

Dan Levi received his B.Sc. degree (with honor) in mathematics and computer science from Tel-Aviv University, in 2000, and his M.Sc. and Ph.D. degrees in applied mathematics and computer science at the Weizmann Institute, Rehovot, Israel, in 2004 and 2009, respectively. In the Weizmann Institute he conducted research in human and computer vision under the instruction of Professor Shimon Ullman. From 2007 he has been conducting industrial computer vision research and development at several companies including Elbit Systems, Israel. Since 2009 he is a researcher at the Smart Sensing and Vision Systems Group at General Motors, Advanced Technical Center-Israel.

Guy Raz received his B.Sc in physics and mathematics from the Hebrew university of Jerusalem (Talpiot program), and his M.Sc and Ph.D. in physics from the Weizmann Institute of Science. For more than 10 years he has been involved in research and development of a large variety of sensing related technologies and systems, with a significant emphasis of electro-optics systems, both at IDF and at Elbit Systems. Currently he is a senior researcher at the Smart Sensing and Vision Systems Group at General Motors, Advanced Technical Center-Israel, working on advanced automotive sensing systems.