Image and Vision Computing 21 (2003) 359–381
www.elsevier.com/locate/imavis

A survey of video processing techniques for traffic applications

V. Kastrinaki, M. Zervakis*, K. Kalaitzakis

Digital Image and Signal Processing Laboratory, Department of Electronics and Computer Engineering, Technical University of Crete, Chania 73100, Greece

Received 29 October 2001; received in revised form 18 December 2002; accepted 15 January 2003

Abstract

Video sensors become particularly important in traffic applications mainly due to their fast response, easy installation, operation and maintenance, and their ability to monitor wide areas. Research in several fields of traffic applications has resulted in a wealth of video processing and analysis methods. Two of the most demanding and widely studied applications relate to traffic monitoring and automatic vehicle guidance. In general, systems developed for these areas must integrate, amongst their other tasks, the analysis of their static environment (automatic lane finding) and the detection of static or moving obstacles (object detection) within their space of interest. In this paper we present an overview of image processing and analysis tools used in these applications and we relate these tools with complete systems developed for specific traffic applications. More specifically, we categorize processing methods based on the intrinsic organization of their input data (feature-driven, area-driven, or model-based) and the domain of processing (spatial/frame or temporal/video). Furthermore, we discriminate between the cases of static and mobile camera. Based on this categorization of processing tools, we present representative systems that have been deployed for operation. Thus, the purpose of the paper is threefold. First, to classify image-processing methods used in traffic applications. Second, to provide the advantages and disadvantages of these algorithms. Third, from this integrated consideration, to attempt an evaluation of shortcomings and general needs in this field of active research.

Keywords: Traffic monitoring; Automatic vehicle guidance; Automatic lane finding; Object detection; Dynamic scene analysis

1. Introduction

The application of image processing and computer vision techniques to the analysis of video sequences of traffic flow offers considerable improvements over the existing methods of traffic data collection and road traffic monitoring. Other methods including the inductive loop, the sonar and microwave detectors suffer from serious drawbacks in that they are expensive to install and maintain and they are unable to detect slow or stationary vehicles. Video sensors offer a relatively low installation cost with little traffic disruption during maintenance. Furthermore, they provide wide area monitoring allowing analysis of traffic flows and turning movements (important to junction design), speed measurement, multiple-point vehicle counts, vehicle classification and highway state assessment (e.g. congestion or incident detection) [1].

Image processing also finds extensive applications in the related field of autonomous vehicle guidance, mainly for determining the vehicle's relative position in the lane and for obstacle detection. The problem of autonomous vehicle guidance involves solving different problems at different abstraction levels. The vision system can aid the accurate localization of the vehicle with respect to its environment, which is composed of the appropriate lane and obstacles or other moving vehicles. Both lane and obstacle detection are based on estimation procedures for recognizing the borders of the lane and determining the path of the vehicle. The estimation is often performed by matching the observations (images) to an assumed road and/or vehicle model.

Video systems for either traffic monitoring or autonomous vehicle guidance normally involve two major tasks of perception: (a) estimation of road geometry and (b) vehicle and obstacle detection. Road traffic monitoring aims at the acquisition and analysis of traffic figures, such as presence and numbers of vehicles, speed distribution data, turning traffic flows at intersections, queue-lengths, space and time occupancy rates, etc. Thus, for traffic monitoring it is essential to detect the lane of the road and then sense

* Corresponding author. Tel.: +30-28210-37206; fax: +30-28210-37542.
E-mail address: michalis@systems.tuc.gr (M. Zervakis).

0262-8856/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0262-8856(03)00004-0

and identify presence and/or motion parameters of a vehicle. Similarly, in autonomous vehicle guidance, the knowledge about road geometry allows a vehicle to follow its route and the detection of road obstacles becomes a necessary and important task for avoiding other vehicles present on the road.

In this paper we focus on video systems considering both areas of road traffic monitoring and automatic vehicle guidance. We attempt a state of the art survey of algorithms and tools for the two major subtasks involved in traffic applications, i.e. the automatic lane finding (estimation of lane and/or central line) and vehicle detection (moving or stationary object/obstacle). With the progress of research in computer vision, it appears that these tasks should be trivial. The reality is not so simple; a vision-based system for such traffic applications must have the features of a short processing time, low processing cost and high reliability [2]. Moreover, the techniques employed must be robust enough to tolerate inaccuracies in the 3D reconstruction of the scene, noise caused by vehicle movement and calibration drifts in the acquisition system. The image acquisition process can be regarded as a perspective transform from the 3D world space to the 2D image space. The inverse transform, which represents a 3D reconstruction of the world from a 2D image, is usually indeterminate (ill-posed problem) because information is lost in the acquisition mapping. Thus, an important task of video systems is to remove the inherent perspective effect from acquired images [3,4]. This task requires additional spatio-temporal information by means of additional sensors (stereo vision or other type sensors) or the analysis of temporal information from a sequence of images. Stereo vision and optical flow methods aid the regularization of the inversion process and help recover scene depth. Some of the lane or object detection problems have been already solved as presented in the next sections. Others, such as the handling of uncertainty and the fusion of information from different sensors, are still open problems as presented in Section 4 that traces the future trends.

In our analysis of video systems we distinguish between two situations. The first one is the case in which a static camera observes a dynamic road scene for the purpose of traffic surveillance. In this case, the static camera generally has a good view of the road objects because of the high position of the camera. Therefore, 2D intensity images may contain enough information for the model-based recognition of road objects. The second situation is the case in which one or more vision sensors are mounted on a mobile vehicle that moves in a dynamic road scene. In this case, the vision sensors may not be in the best position for observing a road scene. Then, it is necessary to correlate video information with sensors that provide the actual state of the vehicle, or to combine multisensory data in order to detect road obstacles efficiently [2].

Both lane and object detection become quite different in the cases of stationary (traffic monitoring) and moving camera (automatic vehicle guidance), conceptually and algorithmically. In traffic monitoring, the lane and the objects (vehicles) have to be detected on the image plane, at the camera coordinates. Alternatively, in vehicle guidance, the lane and the object (obstacle) positions must be located at the actual 3D space. Hence, the two cases, i.e. stationary and moving cameras, require different processing approaches, as illustrated in Sections 2 and 3 of the paper. The techniques used for moving cameras can also be used for stationary cameras. Nevertheless, due to their complexity and computational cost, they are not well suited for the relative simpler applications of stationary video analysis.

Research in the field started as early as in the 70s with the advent of computers and the development of efficient image processing techniques. There is a wealth of methods for either traffic monitoring or terrain monitoring for vehicle guidance. Some of them share common characteristics and some originate from quite diverse approaches. The purpose of this paper is threefold. First, to classify image-processing methods used in traffic applications. Second, to provide the advantages and disadvantages of these algorithms. Third, from this integrated consideration, to attempt an evaluation of shortcomings and general needs in this field of active research. The paper proceeds by considering the problem of automatic lane finding in Section 2 and that of vehicle detection in Section 3, respectively. In Section 4 we provide a critical comparison and relate processing algorithms with complete systems developed for specific traffic applications. The paper concludes by projecting future trends and developments motivated by the demands of the field and the shortcomings of the available tools.

2. Automatic lane finding

2.1. Stationary camera

A critical objective in the development of a road monitoring system based upon image analysis is adaptability. The ability of the system to react to a changing scene while carrying out a variety of goals is a key issue in designing replacements to the existing methods of traffic data collection. This adaptability can only be brought about by a generalized approach to the problem which incorporates little or no a priori knowledge of the analyzed scene. Such a system will be able to adapt to ‘changing circumstances’, which may include the following: changing light levels, i.e. night–day, or sunny–cloudy; deliberately altered camera scene, perhaps altered remotely by an operator; accidentally altered camera position, i.e. buffeting by the wind or knocks due to foreign bodies; changing analysis goals, i.e. traffic flow to counting or occupancy measurement. Moreover, an adaptive system would ease installation of the equipment due to its ability for self-initialization [1]. Automatic lane finding (ALF) is an important task for an adaptive traffic monitoring system.

ALF can assist and simplify the installation of a detection system. It enables the system to adapt to different environmental conditions and camera viewing positions. It also enables applications in active vision systems, where the camera viewing angle and the focal length of the camera lens may be controlled by the system operator to find an optimum view [5].

The aspects that characterize a traffic lane are its visual difference from the environment and the relatively dense motion of vehicles along the lane. Thus, features that can be easily inferred are the lane characteristics themselves (lane markings and/or road edges) and the continuous change of the scene along the lane area. Based on these features, we can distinguish two classes of approaches in lane detection, namely lane-region detection and lane-border detection (lane markings and road edges). The first class relates the detection of the lane with the changing intensity distribution along the region of a lane, whereas the second class considers directly the spatial detection of lane characteristics. It should be emphasized that the first class considers just changes in the gray-scale values within an image sequence, without addressing the problem of motion estimation. The second class can be further separated, based on the method of describing the lane characteristics. Two general subclasses involve *model-driven* approaches in which deformable templates are iteratively modified to match the road edges, and *feature-driven* approaches in which lane features are extracted, localized and combined to meaningful characteristics. The latter approach limits the computation-intensive processing of images to simply extracting features of interest.

2.2. Moving camera

In the case of automatic vehicle guidance, the lane detection process is designed to (a) provide estimates for the position and orientation of the car within the lane and (b) infer a reference system for locating other vehicles or obstacles in the path of that vehicle. In general, both tasks require two major estimation procedures, one regarding the recognition of the borders of the lane and the second for the prediction of the path of the vehicle. The derivation of the path of the vehicle requires temporal information concerning the vehicle motion, as well as modeling of the state of the car (dynamics and kinematics). Alternatively, the lane recognition task can be based on spatial visual information, at least for the short-range estimation of the lane position. Although some systems have been designed to work on completely unstructured roads and terrain, lane detection has generally been reduced to the localization of specific features, such as lane markings painted on the road surface. Certain assumptions facilitate the lane detection task and/or speed-up the processing [6]:

- Instead of processing entire images, a computer vision system can analyze specific regions (the 'focus of attention') to identify and extract the features of interest.
- The system can assume a fixed or smoothly varying lane width and thereby limit its search to almost-parallel lane markings.
- A system can exploit its knowledge of camera and an assumption of a precise 3D road model (for example, a flat road without bumps) to localize features easier and simplify the mapping between image pixels and their corresponding world coordinates.

Real-time road segmentation is complicated by the great variability of vehicle and environmental conditions. Changing seasons or weather conditions, time of the day, dirt on the road, shadows, spectral reflection when the sun is at low angle and manmade changes (tarmac patches used to repair road segments) complicate the segmentation process. Because of these combined effects, robust segmentation is very demanding. Several features of structured roads, such as color and texture, have been used to distinguish between road and non-road regions in each individual frame. Furthermore, road tracking can facilitate road segmentation based on previous information. This process, however, requires knowledge of the vehicle dynamics, vehicle suspension, performance of the navigation and control systems, etc.

Single-frame analysis has been extensively considered not only in monocular but also in stereo vision systems. The approaches used in stereo vision often involve independent processing on the left and right images and projection of the result to the ground plane through the Helmholtz shear equation, making the assumption of flat road and using piecewise road geometry models (such as clothoids) [7,8]. Furthermore, the inverse perspective mapping can be used to simplify the process of lane detection, similar to the process of object detection considered in Section 3 [4]. The inverse perspective mapping essentially re-projects the two images onto a common plane (the road plane) and provides a single image with common lane structure.

In the case of a moving vehicle, the lane recognition process must be repeated continuously on a sequence of frames. In order to accelerate the lane detection process, there is a need to restrict the computation to a reduced region of interest (ROI). There are two general approaches in this direction. The first restricts the search on the predicted path of the vehicle by defining a search region within a trapezoid on the image plane, which is located through the perspective transform. The second approach defines small search windows located at the expected position of the lane, separated by short spatial distances. A rough prediction of the lane position at subsequent video frames can highly accelerate the lane detection process. In one scheme, the estimated lane borders at the previous frame can be expanded, making the lane virtually wider, so that the actual lane borders at the next frame are searched for within this expanded ROI [9]. In a different scheme, a least squares linear fit is used to extrapolate lane markings and locate the new search windows at the next frames [10].

Following the process of lane detection on the image plane, the result must be mapped on the road (world) coordinate system for navigation purposes. By assuming a flat road model, the distance of a 3D-scene point on the road plane can be readily computed if we know the transformation matrix between the camera and the vehicle coordinate systems. In more general cases, the road geometry has to be estimated in order to derive the transformation matrix between the vehicle and the road coordinate systems. The aspects of relative position estimation are further considered in Section 3, along with the object detection process.

2.3. Automatic lane finding approaches

The fundamental aspects of ALF approaches are considered and reviewed in this section. These approaches are classified into lane-region detection, feature driven and model driven approaches.

2.3.1. Lane-region detection

One method of automatic lane finding with stationary camera can be based upon accumulating a map of significant scene change [5]. The so-called activity map, distinguishes between active areas of the scene where motion is occurring (the road) and inactive areas of no significant motion (e.g. verges, central reservation). To prevent saturation and allow adaptation to changes of the scene, the map generation also incorporates a simple decay mechanism through which previously active areas slowly fade from the map. Once formed, the activity map can be used by a lane finding algorithm to extract the lane positions [1].

The lane-region analysis can be also modeled as a classification problem, which labels image pixels into road and non-road classes based on particular features. A typical classification problem involves the steps of feature extraction, feature decorrelation and reduction, clustering and segmentation. For road segmentation applications, two particular features have been used, namely color and texture [11,12]. In the case of color, the features are defined by the spectral response of the illumination at the red, green and blue bands. At each pixel, the (R,G,B) value defines the feature vector and the classification can be performed directly on the (R,G,B) scatter diagram of the image [12]. The green band contributes very little in the separation of classes in natural scenes and on the (R,B) plane classification can be performed through a linear discriminant function [12], since road pixels cluster nicely, distinct from non-road pixels. The classification process can be based on piece-wise linear discriminant functions, in order to account for varying color conditions on the road (shading, reflections, etc.) [12]. The road segmentation can also be performed using stochastic patter recognition approaches. One can define many classes representing road and/or non-road segments. Each class is represented by its mean and variance of (R,G,B) values and it is a priori likelihood based on the expected number of pixels in each class. Gaussian distributions have been used to model the color classes [11].

The apparent color of an object is not consistent, due to several factors. It depends on the illuminant color, the reflectivity of the object, the illumination and viewing geometry and the sensor parameters. The color of a scene may vary with time, cloud cover and other atmospheric conditions, as well as with the camera position and orientation. Thus, color as a feature for classification requires special treatment and normalization to ensure consistency of the classification results. Once the road has been localized in an image, the color statistics of the road and off-road models need be modified in each class, adapting the process to changing conditions [13]. The hue, saturation, gray-value (HSV) space has been also used as more effective for classification [14].

Besides color, the local texture of the image has been used as a feature for classification [11,15]. The texture of the road is normally smoother than that of the environment, allowing for region separation in its feature space. The texture calculation can be based on the amplitude of the gradient operator at each image area. Ref. [11] uses a normalized gradient measure based on a high-resolution and a low-resolution (smoothed) image, in order to handle shadow interior and boundaries. Texture classification is performed through stochastic patter recognition techniques and unsupervised clustering. Since the road surface is poorly textured and differs significantly from objects (vehicles) and background, grey-level segmentation is likely to discriminate the road surface area from other areas of interest. Unsupervised clustering on the basis of the C-means algorithm or the Kohonnen self-organizing maps can be employed on a 3D input space of features. Two of these features signify the position and the third signifies the grey-level of each pixel under consideration. Thus, the classifier groups together neighboring pixels of similar intensities [16].

The classification step must be succeeded by a region merging procedure, as to combine similar small regions under a single label. Region merging may utilize other sources of information, such as motion. In essence, a map of static regions obtained by simple frame differencing can provide information about the motion activity of neighboring patches candidate for merging [16]. Texture classification can also be effectively combined with color classification based on the confidence of the two classification schemes [11].

2.3.2. Feature-driven approaches

This class of approaches is based on the detection of edges in the image and the organization of edges into meaningful structures (lanes or lane markings) [17]. This class involves, in general, two levels of processing, i.e. feature detection and feature aggregation. The *feature* *detection* part aims at extracting intensity discontinuities. To make the detection more effective, a first step of image enhancement is performed followed by a gradient operator. The dominant edges are extracted based on thresholding of the gradient magnitude and they are refined through thinning operators. At this stage, the direction of edges at each pixel can be computed based on the phase of the gradient and a curvature of line segments can be estimated based on neighborhood relations.

Feature aggregation organizes edge segments into meaningful structures (lane markings) based on short-range or long-range attributes of the lane. Short-range aggregation considers local lane fitting into the edge structure of the image. A realistic assumption that is often used requires that the lane (or the lane marking) width does not change drastically. Hence, meaningful edges of the video image are located at a certain distance apart, in order to fit the lane-width model. Long-range aggregation is based on a line intersection model, based on the assumption of smooth road curvature. Thus gross road boundaries and markings must be directed towards a specific point in the image, the focus of expansion (FOE) of the camera system.

Along these directions, Ref. [4] detects lane markings through a horizontal (linear) edge detector and enhances vertical edges via a morphological operator. For each horizontal line, it then forms correspondences of edge points to a two-lane road model (three lane markings) and identifies the most frequent lane width along the image, through a histogram analysis. All pairs of edge pixels (along each horizontal line) that fall within some limits around this width are considered as lane markings and corresponding points on different scan lines are aggregated together as lines of the road. A similar approach is used in Ref. [18] for auto-calibration of the camera module. The Road Markings Analysis (ROMA) system is based on aggregation of the gradient direction at edge pixels in real-time [19]. To detect edges that are possible markings or road boundaries, it employs a contour following algorithm based on the range of acceptable gradient directions. This range is adapted in real-time to the current state variables of the road model. The system can cope with discontinuities of the road borders and can track road intersections.

Ref. [20] detects brightness discontinuities and retains only long straight lines that point toward the FOE. For each edge point, it preserves the edge direction and the neighboring line curvature and performs a first elimination of edges based on thresholding of the direction and curvature. This is done to preserve only straight lines that point towards the specific direction of the FOE. The feature aggregation is performed through correlation with a synthetic image that encodes the road structure for the specific FOE. The edge detection can be efficiently performed through morphological operators [21–23].

The approach in Ref. [10] operates on search windows located along the estimated position of the lane markings. For each search window, the edges of the lane marking are determined as the locations of maximum positive and negative horizontal changes in illumination. Then, these edge points are aggregated as boundaries of the lane making (paint stripe) based on their spacing, which should approximate the lane-marking width. The detected lanes at near-range are extrapolated to far-range via linear least-squares fit, to provide an estimated lane-marking location for placing the subsequent search windows. The location of the road markings along with the state of the vehicle are used in two different Kalman filters to estimate the near and far-range road geometry ahead of the vehicle [10]. Prior knowledge of the road geometry imposes strong constraints on the likely location and orientation of the lanes. Alternatively, other features have been proposed that capture information about the orientation of the edges, but are not affected drastically by extraneous edges. Along these lines, the LANA algorithm [24] uses frequency-domain features rather than features directly related to the detected edges. These feature vectors are used along with a deformable-template model of the lane markers in a Bayesian estimation setting. The deformable template introduces a priori information, whereas the feature vectors are used to compute the likelihood probability. The parameters of the deformable template are estimated by optimizing the resulting maximum a posteriori objective function [24]. Simpler linear models are used in Ref. [14] for road boundaries and lane markings, with their parameters estimated via a recursive least squares (RLS) filter fit on candidate edge points.

In general, feature driven approaches are highly dependent on the methods used to extract features and they suffer from noise effects and irrelevant feature structures. Often in practice the strongest edges are not the road edges, so that the detected edges do not necessarily fit a straight-line or a smoothly varying model. Shadow edges can appear quite strong, highly affecting the line tracking approach.

2.3.3. Model-driven approaches

In model-driven approaches the aim is to match a deformable template defining some scene characteristic to the observed image, so as to derive the parameters of the model that match the observations. The pavement edges and lane markings are often approximated by circular arcs on a flat-ground plane. More flexible approaches have been considered in Refs. [25,26] using snakes and splines to model road segments. In contrast to other deformable line models, Ref. [26] uses a spline-based model that describes the perspective effect of parallel lines, considering simultaneously both-side borders of the road lane. For small to moderate curvatures, a circular arc is approximated by a second-order parabola, whose parameters must be estimated. The estimation can be performed on the image plane [27] or on the ground plane [24] after the appropriate perspective mapping. Bayesian optimization procedures are often used for the estimation of these parameters.

Model-based approaches for lane finding have been extensively employed in stereo vision systems, where the estimation of the 3D structure is also possible. Such approaches assume a parametric model of the lane geometry, and a tracking algorithm estimates the parameters of this model from feature measurements in the left and right images [28]. In Ref. [28] the lane tracker predicts where the lane markers should appear in the current image based on its previous estimates of the lane position. It then extracts possible lane markers from the left and right images. These feature measurements are passed to a robust estimation procedure, which recovers the parameters of the lane along with the orientation and height of the stereo rig with respect to the ground plane. The Helmholtz shear equation is used to verify that candidate lane markers actually lie on the ground plane [28]. The lane markers are modeled as white bars of a particular width against a darker background. Regions in the image that satisfy this intensity profile can be identified through a template matching procedure. In this form, the width of the lane markers in the image changes linearly as a function of the distance from the camera, or the location of the image row considered. Thus, different templates are used at different image locations along the length of the road, in both the left and right images. Once a set of candidate lane markers has been recovered, the lane tracker applies a robust fitting procedure using the Hough transform, to find the set of model parameters which best match the observed data [28]. A robust fitting strategy is absolutely essential in traffic applications, because on real highway traffic scenes the feature extraction procedure almost always returns a number of extraneous features that are not part of the lane structure. These extra features can come from a variety of sources, other vehicles on the highway, shadows or cracks in the roadway etc.

Another class of model-driven approaches involves the stochastic modeling of lane parameters and the use of Bayesian inference to match a road model to the observed scene. The position and configuration of the road, for instance, can be considered as variables to be inferred from the observation and the a posteriori probability conditioned on this observation [25,29]. This requires the description of the road using small segments and the derivation of probability distributions for the relative positions of these segments on regular road scenes (prior distribution on road geometry). Moreover, it requires the specification of probability distributions for observed segments, obtained using an edge detector on the observed image, conditioned on the possible positions of the road segments (a posteriori distribution of segments). Such distributions can be derived from test data [29].

The 3D model of the road can also be used in modeling the road parameters through differential equations that relate motion with spatial changes. Such approaches using state-variable estimation (Kalman filtering) are developed in Refs. [30,31]. The road model consists of skeletal lines pieced together from clothoids (i.e. arcs with constant curvature change over their run length). The road assumptions define a general highway scene, where the ground plane is flat, the road boundaries are parallel with constant width, the horizontal road curvature changes slowly (almost linearly) and the vertical curvature is insignificant. Assuming slow speed changes, or piecewise constant speed, the temporal change of curvature is linearly related to the speed of the vehicle. Thus, the curvature parameters and their association with the ego-motion of the camera can be formulated into a compact system of differential equations, providing a dynamic model for these parameters. The location of the road boundaries in the image is determined by three state variables, i.e. the vehicle lateral offset from the lane center, the camera heading relative to the road direction, and the horizontal road curvature. The Kalman filtering algorithm is employed in Ref. [32] to estimate the state-variables of the road and reconstruct the 3D location of the road boundaries.

The previous model assumes no vertical curvature and no vertical deviation of the camera with respect to the road. These assumptions imply a flat-road geometry model, which is of limited use in practice. Other rigorous models, such as the hill-and-dale and the zero-bank models have been considered for road geometry reconstruction [12,33]. The hill-and-dale model uses the flat-road model for the two roadway points closest to the vehicle in the image, and forces the road model to move up or down from the flat-road plane so as to retain a constant road width. The zero-bank assumption models the road as a space ribbon generated by a central line-spine and horizontal line-segments of constant width cutting the spine at their midpoint at a normal to the spine's 3D direction. Even more unstructured road geometry is studied in Ref. [34], where all local road parameters are involved in the state-variable estimation process.

Model-driven approaches provide powerful means for the analysis of road edges and markings. However, the use of a model has certain drawbacks, such as the difficulty in choosing and maintaining an appropriate model for the road structure, the inefficiency in matching complex road structures and the high computational complexity.

3. Object detection

3.1. Stationary camera

In road traffic monitoring, the video acquisition cameras are stationary. They are placed on posts above the ground to obtain optimal view of the road and the passing vehicles. In automatic vehicle guidance, the cameras are moving with the vehicle. In these applications it is essential to analyze the dynamic change of the environment and its contents, as well as the dynamic change of the camera itself. Thus, object detection from a stationary camera is simpler in that it involves fewer estimation procedures.

Initial approaches in this field involve spatial, temporal and spatio-temporal analysis of video sequences. Using a sequence of images the detection principle is based essentially on the fact that the objects to be searched for are in motion. These methods prioritize temporal characteristics compared with spatial characteristics, i.e. the detection deals mainly with the analysis of variations in time of one and the same pixel rather than with the information given by the environment of a pixel in one image [35]. More advanced and effective approaches consider object modeling and tracking using state-space estimation procedures for matching the model to the observations and for estimating the next state of the object. The most common techniques, i.e. analysis of the optical flow field and processing of stereo images, involve processing two or more images. With optical-flow-field analysis, multiple images are acquired at different times [36]; stereo images, of course, are acquired simultaneously from different points of view [37]. Optical-flow-based techniques detect obstacles indirectly by analyzing the velocity field. Stereo image techniques identify the correspondences between pixels in the different images. Stereovision has advantages in that it can detect obstacles directly and, unlike optical-flow-field analysis, is not constrained by speed. Several approaches considering different aspects of object and motion perception from a stationery camera are considered in Section 3.3.

3.2. Moving camera

Autonomous vehicle guidance requires the solution of different problems at different abstraction levels. The vision system can aid the accurate localization of the vehicle with respect to its environment, by means of matching observations (acquired images) over time, or matching a single observation to a road model or even matching a sequence of observations to a dynamic model. We can identify two major problems with the efficient recognition of the road environment, namely the restricted processing time for real-time applications and the limited amount of information from the environment. For efficient processing we need to limit the ROI within each frame and process only relevant features within this ROI instead of the entire image. Since the scene in traffic applications does not change drastically, the prediction of the ROI from previously processed frames become of paramount importance. Several efficient methods presented in the following are based on dynamic scene prediction using motion and road models. The problem of limited amount of information in each frame stems from the fact that each frame represents a non-invertible projection of the dynamically changing 3D world onto the camera plane. Since single frames encode only partial information, which could be easily misinterpreted, the systems for autonomous vehicle guidance require additional information in the form of a knowledge-base that models the 3D environment and its changes (self/ego motion or relative motion of other objects). It is possible from monocular vision to extract certain 3D information from a single 2D-projection image, using visual cues and a priori knowledge about the scene. In such systems, obstacle determination is limited to the localization of vehicles by means of a search for specific patterns, possibly supported by other features such as shape, symmetry, or the use of a bounding box [38–40]. Essentially, forward projection of 3D models and matching with 2D observations is used to derive the structure and location of obstacles. True 3D modeling, however, is not possible with monocular vision and single frame analysis.

The availability of only partial information in 2D images necessitates the use of robust approaches able to infer a complete scene representation from only partial representations. This problem concerns the matching of a low-abstraction image to a high-abstraction and complexity object. In other words, one must handle differences between the representation of the acquired data and the projected representation of the models to be recognized. A priori knowledge is necessary in order to bridge the gap between these two representations [41]. A first source of additional information is the temporal evolution of the observed image, which enables the tracking of features over time. Furthermore, the joint consideration of a frame sequence provides meaningful constraints of spatial features over time or vice versa. For instance, Ref. [42] employs smoothness constraints on the motion vectors, which are imposed by the gray-scale spatial distribution. Such form of constraints convey the realistic assumption that compact objects should preserve smoothly varying displacement vectors. The initial form of integrated spatio-temporal analysis operates on a so-called $2\frac{1}{2}$D feature space, where 2D features are tracked in time. Additional constraints can be imposed through the consideration of 3D models for the construction of the environment (full 3D space reconstruction) and the matching of 2D data (observations) with the 3D representation of these models, or their projection on the camera coordinates (pose estimation problem). Such model information, by itself, enables the consideration and matching of relative object poses [43].

With the latest advances in computer architecture and hardware, it becomes possible to consider even the dynamic modeling of 3D objects. This possibility paved the way to fully integrated spatio-temporal processing, where two general directions have been proposed. The first one considers the dynamic matching of low-abstraction (2D image-level) features between the data and the model. Although it keeps continuous track of changes in the 3D model using both road and motion modeling (features in a $3\frac{1}{2}$D space), it propagates the current 2D representation of the model in accordance with the current state of the camera with respect to the road [44]. Thus, it matches the observations with the expected projection of the world onto the camera system and propagates the error for correcting the current (model) hypothesis [31]. The second approach uses a full 4D model, where objects are treated as 3D motion processes in space and time. Geometric shape descriptors together with generic models for motion form the basis for this integrated (4D or dynamic vision) analysis [45]. Based on this representation one can search for features in the 4D-space [45], or can match observations (possibly from different sensors or information sources) and models at different abstraction levels (or projections) [41]. This evolution of techniques and their abilities is summarized in Table 1 that is further discussed in the conclusion, after the consideration of established approaches. Some relevant approaches for moving object detection from a moving camera are summarized in Section 3.3.

3.3. Object detection approaches

Some fundamental issues of object detection are considered and reviewed in this section. Approaches have been categorized according to the method used to isolate the object from the background on a single frame or a sequence of frames.

3.3.1. Thresholding

This is one of the simplest, but less effective techniques, which operates on still images. It is based on the notion that vehicles are compact objects having different intensity form their background. Thus, by thresholding intensities in small regions we can separate the vehicle from the background. This approach depends heavily on the threshold used, which must be selected appropriately for a certain vehicle and its background. Adaptive thresholding can be used to account for lighting changes, but cannot avoid the false detection of shadows or missed detection of parts of the vehicle with similar intensities as its environment [46]. To aid the thresholding process, binary mathematical morphology can be used to aggregate close pixels into a unified object [47]. Furthermore, gray-scale morphological operators have been proposed for object detection and identification that are insensitive to lighting variation [48].

3.3.2. Multigrid identification of regions of interest

A method of directing attention to regions of interest based on multiresolution images is developed in Ref. [5]. This method first generates a hierarchy of images at different resolutions. Subsequently, a region search begins at the top level (coarse to fine). Compact objects that differ from their background remain distinguishable in the low-resolution image, whereas noise and small intensity variations tend to disappear at this level. Thus, the low-resolution image can immediately direct attention to the pixels that correspond to such objects in the initial image. Each pixel of interest is selected according to some interest function which may be a function of the intensity values of its adjacent pixels, edge strength, or successive frame differencing for motion analysis [5].

3.3.3. Edge-based detection (spatial differentiation)

Approaches in this class are based on the edge-features of objects. They can be applied to single images to detect the edge structure of even still vehicles [49]. Morphological edge-detection schemes have been extensively applied, since they exhibit superior performance [4,18,50]. In traffic scenes, the results of an edge detector generally highlight vehicles as complex groups of edges, whereas road areas yield relatively low edge content. Thus the presence of vehicles may be detected by the edge complexity within the road area, which can be quantified through analysis of the histogram [51].

Alternatively, the edges can be grouped together to form the vehicle's boundary. Towards this direction, the algorithm must identify relevant features (often line segments) and define a grouping strategy that allows the identification of feature sets, each of which may correspond to an object of interest (e.g. potential vehicle or road obstacle). Vertical edges are more likely to form dominant line segments corresponding to the vertical boundaries of the profile of a road obstacle. Moreover, a dominant line segment of a vehicle must have other line segments in its neighborhood that are detected in nearly perpendicular directions. Thus, the detection of vehicles and/or obstacles can simply consist of finding the rectangles that enclose the dominant line segments and their neighbors in the image plane [2,30]. To improve the shape of object regions Ref. [52,53] employ the Hought transform to extract consistent contour lines and morphological operations to restore small breaks on the detected contours. Symmetry provides an additional useful feature for relating these line segments, since vehicle rears are generally contour and region-symmetric about a vertical central line [54].

Edge-based vehicle detection is often more effective than other background removal or thresholding approaches, since the edge information remains significant even in variations of ambient lighting [55].

3.3.4. Space signature

In this detection method, the objects to be identified (vehicles) are described by their characteristics (forms, dimensions, luminosity), which allow identification in their environment [56,57]. Ref. [57] employs a logistic regression approach using characteristics extracted from the vehicle signature, in order to detect the vehicle from its background. Alternatively, the space signatures are defined in Ref. [58] by means of the vehicle outlines projected from a certain number of positions (poses) on the image plane from a certain geometrical vehicle model. A camera model is employed to project the 3D object model onto the camera coordinates at each expected position. Then, the linear edge segments on each observed image are matched to the model by evaluating the presence of attributes of an outline, for

Table 1
Progressive use of information in different levels of system complexity and functionality

Information Required / Feature Domain	Frames	Video and/or Stereo	3D Models	Recursive Estimation of Problem Dynamics	Spatio-Temporal Modeling of Scene Process	Processing Tools
<u>1st level</u>: 2D Spatial Features from Frames	Lane detection Traffic density					-Edges/Line -Color and Texture -Thresholding -Background Subtraction
<u>2nd level</u>: 2 ½ D 3D Features Mapped on Camera Plane	Lane detection Moving Object detection	Location Estimation Obstacle Detection	-Projection on 2D (pose detection) -Geometric Model Matching -Identification of special vehicle types			-Motion Field -Motion Parallax -Perspective Mapping -Accumulation of Activity Map -2D Point Pattern Matching
<u>3rd level</u>: 3D Scene Reconstruction and (spatial) Feature Extraction			Vehicle Identification			-3D Model Matching
<u>4th level</u>: 3 ½ D 3D Features Tracked in Time				Lane Following Vehicle Following		-Track Dynamics of 3D Models but only Match 2D -Projections with Observations -Point Pattern Matching in Time
<u>5th level</u>: 4D Integrated Spatio-temporal Analysis				Autonomous Vehicle Guidance		-Model Shape and Motion at once as a single Stochastic Process
<u>6th level</u> Sensor and Data Fusion			Improved 3D Model/Dynamics Representation and Matching			-Sequential Operation -Fusion at Different Abstraction Levels using Knowledge Base Rules -Fuzzy Modeling -Theory of Evidence

each of the pre-established object positions (poses). In a similar framework, Ref. [59] projects the 3D model at different poses to sparse 2D arrays, essentially encoding information about the projected edges. These arrays are used for matching with the image data.

Space signatures can also be identified in an image through correlation or template matching techniques, using directly the typical gray-scale signature of vehicles [60]. Due to the inflexible nature of template matching, a specific template must be created for each type of vehicle to be recognized. This creates a problem, since there are many geometrical shapes for vehicles contained in the same vehicle-class. Moreover, the template mask assumes that there is little change in the intensity signature of vehicles. In practice, however, changes in ambient lighting, shadows, occlusion, and severe light reflection on the vehicle body panels generate serious variation in the spatial signatures of same-type vehicles. To overcome such problems, the TRIP II system [58,61] employs neural networks for recalling space signatures, and exploits their ability to interpolate among different known shapes [62].

Despite its inefficiencies, vehicle detection based on sign patterns does not require high computational effort. Moreover, it enables the system to deal with the tracking process and keep the vehicle in track by continuously sensing its sign pattern in real time.

3.3.5. Background frame differencing

In the preceding methods, the image of motionless objects (background image) is insignificant. On the contrary, this method is based on forming a precise background image and using it for separating moving objects from their background. The background image is specified either manually, by taking an image without vehicles, or is detected in real-time by forming a mathematical or exponential average of successive images. The detection is then achieved by means of subtracting the reference image from the current image. Thresholding is performed in order to obtain presence/absence information of an object in motion [5,35,38].

The background can change significantly with shadows cast by buildings and clouds, or simply due to changes in lighting conditions. With these changing environmental conditions, the background frame is required to be updated regularly. There are several background updating techniques. The most commonly used are averaging and selective updating. In averaging, the background is built gradually by taking the average of the previous background with the current frame. If we form a weighted average between the previous background and the current frame, the background is build through exponential updating [63]. In selective updating, the background is replaced by the current frame only at regions with no motion detected; where the difference between the current and the previous frames is smaller than a threshold [63]. Selective updating can be performed in a more robust averaging form, where the stationary regions of the background are replaced by the average of the current frame and the previous background [50].

3.3.6. Inter-frame differencing

This is the most direct method for making immobile objects disappear and preserving only the traces of objects in motion between two successive frames. The immediate consequence is that stationary or slow-moving objects are not detected. The inter-frame difference succeeds in detecting motion when temporal changes are evident. However, it fails when the moving objects are not sufficiently textured and preserve uniform regions with the background. To overcome this problem, the inter-frame difference is described using a statistical framework often employing spatial Markov random fields [64–66]. Alternatively, in Ref. [64] the inter-frame difference is modeled trough a two-component mixture density. The two components are zero mean corresponding to the static (background) and changing (moving object) parts of the image. Inter-frame differencing provides a crude but simple tool for estimating moving regions. This process can be complemented with background frame differencing to improve the estimation accuracy [67]. The resulting mask of moving regions can be further refined with color segmentation [68] or accurate motion estimation by means of optical flow estimation and optimization of the displaced frame difference [16,67], in order to refine the segmentation of moving objects.

3.3.7. Time signature

This method encodes the intensity profile of a moving vehicle as a function of time. The profile is computed at several positions on the road as the average intensity of pixels within a small window located at each measurement point. The analysis of the time signature recorded on these points is used to derive the presence or absence of vehicles [69]. The time signal of light intensity on each point is analyzed by means of a model with pre-recorded and periodically updated characteristics. Spatial correlation of time signatures allows further reinforcement of detection. In fact, the joint consideration of spatial and time signatures provides valuable information for both object detection and tracking. Through this consideration, the one task can benefit from the results of the other in terms of reducing the overall computational complexity and increasing the robustness of analysis [70]. Along these lines, the adaptable time delay neural network developed for the Urban Traffic Assistant (UTA) system is designed and trained for processing complete image sequences [71]. The network is applied for the detection of general obstacles in the course of the UTA vehicle.

3.3.8. Feature aggregation and object tracking

These techniques can operate on the feature space to either identify an object, or track characteristic points of the object [32]. They are often used in object detection to improve the robustness and reliability of detection and reduce false detection rates. The aggregation step handles features previously detected, in order to find the vehicles themselves or the vehicle queues (in case of congestion). The features are aggregated with respect to the vehicle's geometrical characteristics. Therefore, this operation can be interpreted as a pattern recognition task. Two general approaches have been employed for feature aggregation, namely motion-based and model-based approaches [64]. Motion-based approaches group together visual motion consistencies over time [64,72,73]. Motion estimation is only performed at distinguishable points, such as corners [72,74], or along contours of segmented objects [75], or within segmented regions of similar texture [14,67,70]. Line segments or points can also be tracked in the 3D space by estimating their 3D displacements via a Kalman filter designed for depth estimation [18,64,72,73]. Model-based approaches match the representations of objects within the image sequence to 3D models or their 2D projections from different directions (poses) [44,73]. Several model-based approaches have been proposed employing simple 2D region models (mainly rectangles), active contours and polygonal approximations for the contour of the object, 3D models that can be tracked in time and 4D models for full spatio-temporal representation of the object [73,76].

Following the detection of features, the objects are tracked. Two alternative methods of tracking are employed in Ref. [32], namely numeric signature tracking and symbolic tracking. In signature tracking, a set of intensity and geometry-based signature features are extracted for each detected object. These features are correlated in the next frame to update the location of the objects. Next, the signatures are updated to accommodate for changes in range, perspective, and occlusion. In general, features for tracking encode boundary (edge based) or region (object motion, texture or shape) properties of the tracked object. Active contours, such as snakes and geodesic contours are often employed for the description of boundaries and their evolution over the sequence of frames. For region-based features tracking is based on correspondences among the associated target regions at different time instances [64,77]. In symbolic tracking, objects are independently detected in each frame. A symbolic correspondence is made between the sets of objects detected in a frame pair. A time-sequenced trajectory of each matched object provides a track of the object [32].

3.3.9. Optical flow field

Approaches in this class exploit the fact that the appearance of a rigid object changes little during motion, whereas the drastic changes occur at regions where the object moves in and/or out of the background. The optical flow field $\mathbf{u}(\mathbf{x}, t)$ is computed by mapping the gray-value $g(\mathbf{x} - \mathbf{u}\Delta t, t - \Delta t)$ recorded at time $t - \boldsymbol{\Delta} t$ at the image point $\mathbf{x} - \mathbf{u}\Delta t$ onto the gray-value $g(\mathbf{x}, t)$ recorded at location $\mathbf{x}$ at time t. The optical flow field encodes the temporal displacement of observable gray-scale structures within an image sequence. It comprises information not only about the relative displacement of pixels, but also about the spatial structure of the scene.

Various approaches have been proposed for the efficient estimation of optical flow field [42,78–80]. In general, they can be characterized as (i) gradient-based (ii) correlation based (iii) feature-based and (iv) multigrid methods. Gradient-based techniques focus on matching $g(\mathbf{x} - \mathbf{u}\Delta t, \times t - \Delta t)$ with $g(\mathbf{x}, t)$ on a pixel-by-pixel basis through the temporal gradient of the image sequence. In most cases, the intensity variations alone do not provide sufficient information to completely determine both components (magnitude and direction) of the optical flow field $\mathbf{u}(\mathbf{x}, t)$ [81]. Smoothness constraints facilitate the estimation of optical flow fields even for areas with constant or linearly distributed intensities [78–80,82]. Gradient-based techniques yield poor results for poor-texture images and in presence of shocks and vibrations [83]. Under such conditions, correlation-based techniques usually derive more accurate results. Correlation-based techniques search for the maximum shift around each pixel that maximizes the correlation of gray-level patterns between two consecutive frames. Such procedures are quite expensive in terms of computational complexity. Attempts to speed up the computation at the cost of resolution often imply subsampling of the image and computation of the motion field at fewer image points [83].

Feature-based approaches consider the organization (clustering) of pixels into crude object structures in each frame and subsequently compute motion vectors by matching these structures in the sequence of frames. A robust feature-based method for the estimation of optical flow vectors has been developed by Kories and Zimmermann [84]. Each frame is first subjected to a bandpass filter. Blobs representing local maxima and minima of the gray-level are identified as features. The centroids of the detected blobs are tracked through subsequent frames, resulting in optical flow vectors. A related technique is considered in Ref. [85], which aims at matching areas of similar intensities in two consecutive frames. To reduce the amount of computation, pixels of interest are segmented prior to matching using background removal, edge detection or inter-frame difference. The accuracy of these techniques is affected by sensor noise (quantization), algorithmic disturbances and, more importantly, perspective distortions and occlusion resulting from typical camera positions. Nevertheless, the methods are suitable for on-line qualitative monitoring, operating at much faster speeds than human operators and without the problem of limited attention spans [85].

Multigrid methods are designed for fast estimation of the relevant motion vectors at low resolution and hierarchical refinement of the motion flow field at higher resolution levels [86]. The multigrid approach in Ref. [5] relies upon the organization of similar pixel-intensities into objects, similar to the feature based approaches. This approach, however, identifies object structures at low-resolution levels where it also computes a crude estimate of the motion field from the low-resolution image sequence. The motion vector field is refined hierarchically at higher resolution levels. A related approach is used in the ACTIONS system, where the optical flow vectors are clustered in order to incrementally create candidate moving-objects in the picture domain [81].

For a still camera, moving objects are readily identified by thresholding the optical flow field. The detection of moving objects in image sequences taken from a moving camera becomes much more difficult due to the camera motion. If a camera is translating through a stationary environment, then the directions of all optical-flow vectors intersect at one point in the image plane, the focus of expansion or the epipole [81]. When the car bearing the camera is moving in a stationary environment along a flat road and the camera axis is parallel to the ground, the motion field (due to ego-motion) is expected to have almost quadratic structure [83]. If another moving object becomes visible by the translating camera, the optical flow field resulting from this additional motion will interfere with the optical flow field of the ego-motion. This interference can be detected by testing if the calculated optical-flow vectors have the same direction as the estimated ego-motion model vectors [81,83]. The detection of obstacles from a moving camera based on the optical flow field is generally divided into two steps. The ego-motion is first computed from the analysis of the optical flow. Then, moving or stationary obstacles are detected by analyzing the difference between the expected and the real velocity fields [36,72,87]. These fields are re-projected to the 3D road coordinate system using a model of the road (usually flat straight road) [88,89].

The estimation of ego-motion can be based on parametric models of the motion field. For planar motion with no parallax (no significant depth variations), at most eight parameters can characterize the motion field. These parameters can be estimated by optimizing an error measure on two subsequent frames using a gradient-based estimation approach [66,90]. The optimization process is often applied on a multiresolution representation of the frames, to provide robust performance of the algorithm [90]. When the scene is piecewise planar, or is composed of a few distinct portions at different depths, then the ego-motion can be estimated in layers of 2D parametric motion estimation. Each layer estimates motion at a certain depth due to the camera and removes the associated portions of the image. Image regions that cannot be aligned in two frames at any depth are segmented into independently moving objects [90]. For more general motion of the camera, the ego-motion effect can be decomposed into the planar and the parallax parts. After compensating for the planar 2D motion, the residual parallax displacements in two subsequent frames are primarily due to translational motion of the camera. These displacements due to camera motion form a radial field centered at the epipole. Independently moving objects can be recovered by verifying that the displacement at any given point is directed away from the epipole [91].

The problem of recovering the optical flow from time-varying image sequences is ill-posed and additional constraints must be often imposed to derive satisfactory solutions. Smoothness constraints stem from the fact that uniformly moving objects possess slightly changing motion fields. Such constraints have been used in a joint spatio-temporal domain of analysis [92]. Ref. [93] first calculates the optical flow and after smoothing the displacement vectors in both the temporal and the spatial domains, it merges regions of relatively uniform optical flow. Finally, it employs a voting process over time in each spatial location regarding the direction of the displacement vectors to derive consistent trends in the evolution of the optical flow field and, thus, define consistently moving objects. In a different form, Ref. [94] starts from similarity in the spatial domain. For each frame, it defines characteristic features (such as corners and edges) and matches these features on the present and the previous frame to derive a list of flow vectors. Similar flow vectors are grouped together and compared to the spatial features, in order to verify not only temporal but also spatial consistency of detected moving objects. In a similar form, Ref. [94] defines patches of similar spatial characteristics in each frame and uses local voting over the output of a correlation-type motion detector to detect moving objects. It also uses the inverse perspective mapping to eliminate motion effects on the ground plane due to the ego-motion of the camera [94].

3.3.10. Motion parallax

When the camera is moving forward towards an object, the object's projection on the 2D image plane also moves relative to the image coordinate system. If an object extends vertically from the ground plane, its image moves differently from the immediate background. Moreover, the motion of points on the same object appears different relative to the background, depending on the distance from the ground plane. This difference is called motion parallax [87]. If the environment is constrained, e.g. motion on a planar road, then differences observed on the motion-vector can be used to derive information regarding the objects moving within the scene. If we use the displacement field of the road to displace the object, a clear difference between the predicted and the actual position of the object is experienced. In other words, all points in the image that are not on the ground plane will be erroneously predicted. Thus, the prediction error (above an acceptable threshold) indicates locations of vertically extended objects in the scene [87]. If we compensate the ego-motion of the camera, then independently moving (or stationary) obstacles can be readily detected.

The parallax effect is used in the Intelligent Vehicle (IV) in a different form for obstacle detection [95]. A stereo rig is positioned vertically, so that one camera is located above the other. Obstacles located above the ground plane appear identical in the camera images, except from their different location. On the other hand, figures on the road appear different on the two cameras. In this configuration, an obstacle generates the same time signature, whereas road figures generate different time signatures on the two cameras. Thus, progressive scanning and delaying one of the camera signals make the detection of obstacles possible. Nevertheless, the system relies on and is highly affected by brightness changes, shadows and shades on the road structure [95].

3.3.11. Stereo vision

The detection of stationary or moving objects in traffic applications has been also considered through stereo vision systems. The disparity between points in the two stereo images relates directly to the distance of the actual 3D location from the cameras. For all points lying on a plane, the disparity on the two stereo images is a linear function of image coordinates (Helmholtz shear equation). This Helmholtz shear relation highly simplifies the computation of stereo disparity. It may be used to re-map the right image onto the left, or both images onto the road coordinate system, based on the given model of the road in front of the vehicle (e.g. flat straight road) [6,7,28,38,96]. All points on the ground plane appear with zero disparities, whereas residual disparities indicate objects lying above the ground plane and can become potential obstacles. A simple threshold can be used to identify these objects in the difference of the re-mapped images.

Besides the projection of images onto the ground plane, stereo vision can be effectively used for the reconstruction of the 3D space ahead of the vehicle. This reconstruction is based on correspondences between points in the left and right images. Once this has been accomplished, the 3D coordinates of the matched point can be computed via a re-projection transform. The approach in the Path project [28] considers such a matching of structural characteristics (vertical edges). Candidate matches in the left and right images are evaluated by computing the correlation between a window of pixels centered on each edge [28]. The matching can also be based on stochastic modeling, which can take under consideration the spatial intra and inter-correlation of the stereo images [97]. The re-projection transform maps the matched points onto the road coordinate system. For this purpose it is necessary to know the exact relationship among the camera, vehicle and road coordinate systems. Under the assumption of a flat road, this re-projection process is quite straightforward (triangulation transform). In the case of general road conditions, however, the road geometry has to be estimated first in order to derive the re-projection transform from the camera to road coordinate systems. This estimation requires the exact knowledge of the state of the car (yaw rate, vehicle speed, steering angle, etc.), which can be provided by appropriate sensors of the vehicle. Using this information, the road geometry can be estimated from visual data [10,45].

3.3.12. Inverse perspective mapping

A promising approach in real-time object detection from video images is to remove the inherent perspective effect from acquired single or stereo images. The perspective effect relates differently 3D points on the road (world) coordinate system with 2D pixels on the image plane, depending on their distance from the camera. This effect associates different information content to different image pixels. Thus, road markings or objects of the same size appear smaller in the image as they move away from the camera coordinate system. The inverse perspective mapping aims at inverting the perspective effect, forcing homogeneous distribution of information within the image plane. To remove the perspective effect it is essential to know the image acquisition structure with respect to the road coordinates (camera position, orientation, etc.) and the road geometry (the flat-road assumption highly simplifies the problem). The inverse perspective mapping can be applied to stereovision [4], by re-mapping both right and left images into a common (road) domain. Using this approach, the localization of the lane and the detection of generic obstacles on the road can be performed without any 3D-world reconstruction [4]. The difference of the re-mapped views transforms relatively square obstacles into two neighboring triangles corresponding to the vertical boundaries of the object, which can be easily detected on a polar histogram of the difference image.

3.3.13. 3D modeling and forward mapping

The previous approaches reflect attempts to invert the 3D projection for a sequence of images and reconstruct the actual (world) spatial arrangement and motion of objects. The class of model-based techniques takes a different approach. It tries to solve the analysis task by carrying out an iterative synthesis with prediction error feedback using spatio-temporal world models.

Model based approaches employ a parameterized 3D vehicle model for both its structural (shape) characteristics and its motion [73,76]. Considering first a stationery camera, two major problems must be solved, namely the model matching and the motion estimation. The model matching process aims at finding the best match between the observed image and the 3D model projected onto the camera plane. This step is essentially a pose identification process, which derives the 3D position of the vehicle relative to the camera coordinates, based on 2D projections. The vehicle model often assumes straight line segments represented by their length and mid point location [44]. The line segments extracted from the image are matched to the model segments projected on the 2D camera plane. The matching can be based on the optimization of distance measures between the observation and the model; the Mahalanobis distance is used in Ref. [44]. The motion estimation process is based on models that describe the vehicle motion. The motion parameters of this model are estimated using a time-recursive estimation process. For instance, the maximum a posteriori (MAP) estimator is employed in Ref. [44], whereas the extended Kalman filter is used in Ref. [98]. The estimation of motion and shape parameters can be combined in a more general (overall) state estimation process [98].

In the case of a moving camera, the changing views of objects during self or ego-motion reveal different aspects of the 3D geometry of objects and their surrounding environment. It becomes obvious that knowledge about the structure of the environment and the dynamics of motion are relevant components in real-time vision. In a computerized system, generic models of objects from the real world can be stored as three-dimensional structures carrying visible features at different spatial positions relative to their center of gravity. From the ego-motion dynamics, the relative position of the moving vehicle and its camera can be inferred. From this knowledge and applying the laws of forward projection (which is done much more easily than the inverse), the position and orientation of visual features in the image can be matched to those of the projected model [31,34,66]. In a different form, Ref. [31] models the remaining difference image from two consecutive frames after ego-motion compensation as a Markov random field (MRF) that incorporates the stochastic model of the hypothesis that a pixel is either static (background) or mobile (vehicle). The MRF also induces spatial and temporal smoothness constraints. The optimization of the energy function of the resulting Gibbs posterior distribution provides the motion-detection map at every pixel [66].

The dynamic consideration of a world model allows not only the computation of present vehicle positions, but also the computation of the effects of each component of the relative state vector on the vehicle position. This information can be maintained and used for estimating future vehicle positions. The partial derivatives for the parameters of each object at its current spatial position are collected in the Jacobian matrix as detailed information for interpreting the observed image. The ego-motion dynamics can be computed from the actuators of the moving vehicle. The dynamics of other moving obstacles can be modeled by stochastic disturbance variables. For simplicity, the motion of obstacles can be decomposed into translation and rotation over their center of gravity. Having this information, we can proceed with a prediction of the vehicle and obstacles' states for the next time instant, when new measurements are taken. If the cycle time of the measurement and control process is small and the state of the object is well known, the discrepancy between prediction and measurement should be small. Therefore, a linear approximation to the non-linear equations of the model should be sufficient for capturing the essential inter-relationships of the estimation process [31]. Moreover, for linear models the recursive state estimation is efficiently performed through least-square processes. Thus, the spatial state estimation through vision can be performed through recursive least squares estimation and Kalman filtering schemes, where the Jacobian matrix reflects the observed image variation.

By applying this scheme to each object in the environment in parallel, an internal representation of the actual environment can be maintained in the interpretation process, by prediction error feedback [31]. A Kalman filter can be used to predict the vector of the state estimates based on the vectors of measurements and control variables. The measurement equation has to be computed only in the forward direction, from state variables (3D world) to measurement space (image plane). This approach avoids the ill-posed approximation of the non-unique inverse projection transform, through the fusion of dynamic models (that describe spatial motion) with 3D shape models (that describe spatial distribution of visual features). The forward projection mapping is easily evaluated under the flat-road model [31]. Other road models, including the Hill-and-Dale, Zero-Bank and Modified Zero-Bank models have been considered along with the inverse and/or forward mapping [12,33,34].

4. Representative systems and future trends

Based on the previous categorization of video analysis methods, we attempt a brief review of existing systems for traffic monitoring and automatic vehicle guidance. It should be mentioned that this review does not by any means cover all existing systems, but it rather considers representative systems that highlight the major trends in the area. We also attempt a categorization of these systems in terms of their domain of operation, the basic processing techniques used and their major applications. This categorization is summarized in Table 2. More specifically, the fundamental processing techniques from Sections 2 and 3 are summarized for each system. Furthermore, their operating domain is classified in terms of the nature of features utilized. Thus, we consider operation simply in the spatial domain, spatial features with temporal projection of feature locations, temporal features (optical flow field) constrained on spatial characteristics (mainly $2\frac{1}{2}$D and joint spatio-temporal operation. Whenever important, we also emphasize the estimation of vehicle's state variables. In terms of their major applications, we first indicate the status of the camera (static or moving) and categorize applications as in traffic monitoring, automatic lane finding, lane following, vehicle following or autonomous vehicle guidance.

In summary, this paper provides a review of video analysis tools and their operation in traffic applications. The review focuses on two areas, namely automatic lane finding and obstacle detection. It attempts to compile the differences in the requirements and the constraints of these two areas,

Table 2
Representative systems and their functionality

System	Operating domain	Processing techniques	Major applications
ACTIONS [81]	• Spatio-temporal	• Optical flow field with spatial smoothness constraints	• Traffic monitoring • Static camera
AUTOSCOPE [99–101]	• Spatial and temporal domain independently	• Background frame differencing & interframe differencing • Edge detection with spatial and temporal gradients for object detection	• Traffic monitoring • Static camera
CCATS [69]	• Temporal-domain with spatial constraints	• Background removal and model of time signature for object detection	• Traffic monitoring • Static camera
CRESTA [102]	• Temporal-domain differences with spatial constraints	• Interframe differencing for object detection	• Traffic monitoring • Static camera
IDSC [103]	• Spatial domain process with temporal background updating	• Background frame differencing	• Traffic monitoring • Static camera
MORIO [89]	• Spatial domain processing with temporal tracking of features	• Optical flow field • And 3D object modeling	• Traffic monitoring • Static camera
TITAN [47]	• Spatial operation with temporal tracking of features	• Background frame differencing • and Morphological processing for vehicle segmentation	• Traffic monitoring • Static camera
TRIP II [61]	Spatial operation with neural nets	• Spatial signature with neural nets for object detection	• Traffic monitoring • Static camera
TULIP [104]	• Spatial operation	• Thresholding for object detection	• Traffic monitoring • Static camera
TRANSVISION [5]	• Spatio-temporal domain	• Lane region detection (activity map) • and background frame differencing for object detection	• Traffic monitoring • Static camera
VISATRAM [105]	• Spatio-temporal domain	• Background frame differencing • Inverse perspective mapping • Spatial differentiation • Tracking on epipolar plane of the spatio-temporal cube	• Traffic monitoring • Static camera
ARCADE [106]	• Spatial processing	• Model-driven approach with deformable templates for edge matching	• Automatic Lane finding • Static camera
LANELOK [107]	• Spatial processing	• Model-driven approach with deformable templates for edge matching	• Automatic lane finding • Moving camera

(*continued on next page*)

Table 2 (*continued*)

System	Operating domain	Processing techniques	Major applications
LANA [24]	• Spatial processing	• Model-driven approach exploiting features to compute likelihoud • DCT features • Deformable template models for priors	• Automatic lane finding • Moving camera
LOIS [9]	• Spatial processing	• Model-driven approach with deformable templates for edge matching	• Automatic lane finding • Moving camera
CLARK [108]	• Spatial processing of images • Temporal estimation of range observations	• LOIS for lane detection • Color and deformable templates for object detection	• Automatic lane finding and obstacle detection • Moving camera
PVS and AHVS [95]	• Spatial processing	• Feature-driven approach Using edge detection	• Automatic lane finding • Moving camera
RALPH [109]	• Spatial processing • Stereo images	• Feature-driven approach using edge orientation • Mapping of left to right image features	• Automatic lane finding • Moving camera
ROMA [19]	• Spatial operation with tempo-ral projection of lane location	• Feature-driven approach Using edge orientation	• Automatic lane finding • static camera
SIDE WALK [110]	• Spatial processing	• Lane-region detection • via Thresholding for area segmentation	• Automatic lane finding • Moving camera
SCARF [111]	• Spatial processing	• Model-driven approach • Using stochastic modeling for image segmentation	• Automatic lane following • Moving camera
CAPC [10]	• Spatial-domain lane finding with temporal projection of lane location • Temporal estimation of vehicle's state variables	• Feature-driven approach • Using edge detection and constraints on model for lane width and lane spacing	• Automatic lane following • Moving camera
ALV [12]	• Spatial-domain lane and object detection • Temporal estimation of vehicle's state variables	• Lane-region detection for ALF using color classification • Spatial signature for object detection via color segment.	• Automatic lane following • Moving camera
NAVLAB [11]	• Spatial-domain lane finding • Temporal estimation of vehicle's state variables for 3D road-geometry estimation and projection of frame-to-world coordinates	• Lane-region detection for alf via color and texture classification	• Automatic lane following • Moving camera
ALVINN and MANIAC [112]	• Spatial-processing with neural nets	• Recognition of space signature of road through neural nets • Moving camera • Form of temporal matching	• Automatic lane following
Ref. [3]	• Spatial-processing	• Model-driven approach	• Automatic lane following

Table 2 (*continued*)

System	Operating domain	Processing techniques	Major applications
		• Using multiresolution estimation of lane position and orientation	• Moving camera
Ref. [113]	• Spatial-processing with temporal projection of lane location • Temporal estimation of vehicle's state variables	• Model-driven approach • 3D modeling of lane markings and borders	• Automatic lane following • Moving camera
Ref. [41]	• Spatial detection of 2D line segments • Temporal projection of segment locations	• Interframe differencing and edge detection for 2D line segments • Inverse projection of 2D line segments and grouping in 3D space via coplanar transforms	• Vehicle recognition and tracking (pose estimation) • Moving camera
Ref [114]	• Spatial detection of 2D line segments • Temporal projection of segment locations • Temporal estimation of vehicle's state variables for ego-motion estimation	• Spatial signatures of object discontinuities • Inverse projection of 2D features in the 3D space and tracking of 3D features	• Vehicle recognition and tracking • Moving camera
Ref. [115]	• Spatio-temporal processing • Temporal prediction of optical flow • Temporal estimation of vehicle's state variables	• Optical flow field estimation, constrained on spatial edges for object detection	• Obstacle avoidance • Moving camera
BART [116]	• Spatio-temporal processing • Stereo images • Temporal estimation of vehicle's state	• Feature tracking for object detection • Projection of 3D coordinates on 2D stereo images	• Vehicle following • Moving camera
IV (INTELLIGENT VEHICLE) [95]	• Spatio-temporal processing	• Motion parallax for object detection	• Autonomous vehicle guidance • Moving camera
PATH Project [28,117]	• Spatial-domain for alf with temporal projection of lane locations • Spatial correspondence of structure in stereo images for object detection • Stereo images	• Model-driven approach • Temporal matching for lane detection • Hough transform to estimate line model for object edges • Stereo matching of object lines	• Autonomous vehicle guidance • Moving camera
GOLD system [4] for ARGO and MOB-LAB vehicles (Prometheus project)	• Spatial-domain processing for alf and object detection • Temporal projection of lane locations • Temporal estimation of vehicle's state variables	• Feature-driven approach • Edge detection constrained on lane width in each stereo image for alf	• Autonomous vehicle guidance • Moving camera

(*continued on next page*)

Table 2 (*continued*)

System	Operating domain	Processing techniques	Major applications
		• Edge detection individually on each stereo image For object detection • Detection of lane and object through inverse perspective mapping	
VAMORS [31] (Prometheus project)	• Spatio-temporal processing for ALF and vehicle guidance • Temporal estimation of vehicle's state variable	• Model-driven approach • 3D object modeling and forward perspective mapping • State-variable estimation of road skeletal lines for Alf • State-variable estimation of 3D model structure for object detection	• Autonomous vehicle guidance • Moving camera
UTA [14]	• Spatio-temporal processing for ALF and object detection Based on neural networks	• Feature-driven approach • Use of spatio-temporal signature	• Autonomous vehicle guidance • Moving camera
Ref. [14]	• Spatial processing for alf and object detection • Feature tracking in temporal domain	• Feature-driven approach • Color road detection and lane detection via RLS fitting • Interframe differencing and edge detection for locating potential object templates • Feature tracking via RLS	• Autonomous vehicle guidance • Moving camera

which lead to different processing techniques on various levels of information abstraction. Video sensors have demonstrated the ability to obtain traffic measurements more efficiently than other conventional sensors. In cases emulating conventional sensors, video sensors have been shown to offer the following advantages: competitive cost, non-intrusive sensing, lower maintenance and operation costs, lower installation cost and installation/operation during construction. However, because video sensors have the potential of wide area viewing, they are capable of more than merely emulating conventional sensors. Some additional measurements needed for adaptive traffic management are: approach queue length, approach flow profile, ramp queue length, vehicle deceleration, and automatic measurement of turning movements. Vision also provides powerful means for collecting information regarding the environment and its actual state during autonomous locomotion. A vision-based guidance system applied to outdoor navigation usually involves two main tasks of perception, namely *finding the road geometry* and *detecting the road obstacles*. First of all, the knowledge about road geometry allows a vehicle to follow its route. Subsequently, the detection of road obstacles is a necessary and important task to avoid other vehicles present on a road. The complexity of the navigation problem is quite high, since the actual task is to reconstruct an inherent 3D representation of the spatial environment from the observed 2D images.

Image processing and analysis tools become essential components of automated systems in traffic applications, in order to extract useful information from video sensors. From the methods presented in this survey, the big thrust is based on traditional image processing techniques that employ either similarity or edge information to detect roads and vehicles and separate them from their environment. This is evident from the comprehensive overview of systems and their properties in Table 2. Most of these algorithms are rather simple, emphasizing more on high processing speeds than on accuracy of the results. With the rapid progress in the electronics industry, computational complexity becomes less restrictive for real-time applications. Thus, modern hardware systems allow for more sophisticated and accurate algorithms to be employed that capitalize on the real advantages of machine vision. Throughout this work we emphasize on the trend for utilizing more and more information, in order to accurately match the dynamically changing 3D world to the observed image sequences. This evolution is graphically depicted in Table 1, which relates applications to required pieces of information depending on the complexity of each application. The levels of abstraction and information analysis range from 2D frame processing to full 4D spatio-temporal analysis, allowing the development of applications from simple lane detection to the extremely complex task of autonomous vehicle guidance.

A question that arises at this point concerns the future developments in the field. Towards the improvement of the image-processing stage itself, we can expect morphological operators to be used more widely for both the segmentation of smooth structures and the detection of edges. Such non-linear operators provide algorithmic robustness and increased discrimination ability in complex scenes, such as in traffic applications. Furthermore, we can expect increased use of multiresolution techniques that provide not only detailed localized features in the scale-space or in the wavelet domains but also abstract overall information, simulating more closely the human perception.

Most sophisticated image processing approaches adopt the underlying assumption that there exists hard evidence in the measurements (image) to provide characteristic features for classification and further mapping to certain world models. Recursive estimation schemes (employing Kalman filters) proceed in a probabilistic mode that first derives the most likely location of road lanes and vehicles and then matches the measured data to these estimated states. Nevertheless, they also treat the measurements as hard evidence that provide indisputable information. A few approaches deal with difficult visual conditions (shades, shadowing, changing lighting), but they also rely heavily on the measurements and the degree of discrimination that can be inferred from them. Since measurements in real-life are amendable to dispute, it is only natural to expect the development of systems for traffic application that are based on so-called soft computing techniques.

Real-world traffic applications must account for several aspects that accept diverse interpretation, such as the weather, light, static or moving objects on the road, noise, manmade interventions, etc. It is difficult for a vision algorithm to account for all kinds of variations. In fact, it is impossible to define thresholds and other needed parameters for feature extraction and parameter estimation under all different situations. Moreover, under diverse conditions, different algorithms or even different modality systems provide quite different results and/or information regarding the same scene. Thus, two major requirements from advanced systems are expected to emerge namely (i) adaptation to environmental changes and (ii) ability of combining (fusing) information from different sources.

The first issue of adaptability can be dealt either with systems that can be trained in diverse conditions or with systems that tolerate uncertainty. The first class comprises systems that can be trained in some conditions and have the ability to interpolate among learnt situations when an unknown condition is presented. Neural networks form powerful structures for capturing knowledge and interpretation skills by training. In traffic applications there are enough data for extensive training that can be gathered on-line from a fixed sensor (camera) position or while a human operator drives the vehicle. The second class of adaptable systems can deal with more general scenarios, like rainy or extremely hot conditions, where the measurements may not allow for indisputable inference. Under such circumstances, 'illusion' patterns on the scene may be easily misinterpreted as vehicles or roadway structures. In these cases, the measurements convey a large degree of uncertainty. Approaches that deal with uncertainty possibly based on fuzzy-set theory have not been studied in transportation applications. Since they provide powerful means of incorporating possibility and linguistic interpretations expressed by human experts, they are expected to dominate in systems that can intelligently adapt to the environment.

The second requirement for information fusion is still in primitive stages. Due to the complexity of a dynamically changing road scene, vision sensors may not provide enough information to analyze the scene. Then, it is necessary to combine multisensory data in order to detect road characteristics and obstacles efficiently. Representative systems are developed in the OMNI project for traffic management that combines video sensors with information from vehicles equipped with GPS/GSM [18] and in the autonomous vehicle of Ref. [118] that combines vision and laser radar systems with DGPS localization data and maps. The fusion of different sources of information, mainly video and range data has been considered as a means of providing more evidence in reconstructing the 3D world in robotic applications. Nevertheless, most of these approaches use data sources in sequential operation, where the video image guides the focus for attention of the range finder or vice versa [33,108,119,120].

We can distinguish two kinds of multisensory cooperation, the active and intelligent sensing (sequential sensor operation) and the fusion-oriented sensing (simultaneous sensor operation). The former uses the results of 2D image analysis to guide range sensing. Pioneer work with this idea is presented in Refs. [121,122]. For instance, Ref. [121] shows how to recognize 3D objects by first detecting characteristic points in images and then using them to guide the acquisition of 3D information around these points. The second scheme adopts the strategy of combining both intensity and range information to facilitate the perception task [33,123]. Along these lines, Ref. [123] develops and uses a (combined) range-intensity histogram for the purpose of identifying military vehicles. Ref. [124] uses a trainable neural network structure for fusing information from a video sensor and a range finder, in order to determine the appropriate turn curvature so as to keep the vehicle at the middle of the road. Similar strategies can be used for fusing results of different image processing techniques on the same set of data. Individual Processing algorithms provide specific partial solutions under given constraints. The results of such algorithms are not independent, revealing a redundancy of information that can provide robust and reliable solutions if results of individual algorithms are suitably combined or fused. An approach for fusion of algorithms is developed in Ref. [125] using neural networks. Another approach for optimal fusion through Kalman filtering is developed in Ref. [85]. In all these approaches, strict assumptions regarding the coordination of the data sources must be used for the formulation of the fusion algorithm. The fully cooperative fusion of information from different sources that observe the same dynamic scene simultaneously, where evidence from one source can support (increase) or reject (reduce) evidence from the other source, is still in primitive stages and is expected to receive increasing attention. Even less established is the fusion of (probabilistic) hard and (fuzzy) soft evidence [126,127]. The Theory of Evidence provides tools for such a synergetic consideration of different data analysis techniques or data acquisition sensors, where evidence regarding the same spatio-temporal scene from two or more sources is jointly combined to provide increased assurance about the classification results. Towards this direction, the theory of evidence has been presented as a valuable tool for information and/or sensor fusion [128].

References

[1] B.D. Stewart, I. Reading, M.S. Thomson, T.D. Binnie, K.W. Dickinson, C.L. Wan, Adaptive lane finding in road traffic image analysis, Proceedings of Seventh International Conference on Road Traffic Monitoring and Control, IEE, London (1994).

[2] M. Xie, L. Trassoudaine, J. Alizon, J. Gallice, Road obstacle detection and tracking by an active and intelligent sensing strategy, Machine Vision and Applications 7 (1994) 165–177.

[3] A. Broggi, S. Berte, Vision-based road detection in automotive systems: a real-time expectation-driven approach, Journal of Artificial Intelligence Research 3 (1995) 325–348.

[4] M. Bertozzi, A. Broggi, GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection, IEEE Transaction Image Processing 7 (1) (1998).

[5] C.L. Wan, K.W. Dickinson, T.D. Binnie, A cost-effective image sensor system for transport applications utilising a miniature CMOS single chip camera, IFAC Transportation systems, Tianjin, Proceedings (1994).

[6] M. Bertozzi, A. Broggi, Vision-based vehicle guidance, Computer Vision 30 (7) (1997).

[7] J. Weber, D. Koller, Q.-T. Luong, J. Malik, New results in stereo-based automatic vehicle guidance, Proceedings of IEEE Intelligent Vehicles 95, Detroit (1995) 530–535.

[8] Q.-T. Luong, J. Weber, D. Koller, J. Malik, An integrated stereo-based approach to automatic vehicle guidance, Proceedings of the Fifth ICCV, Boston (1995) 12–20.

[9] A. Soumelidis, G. Kovacs, J. Bokor, P. Gaspar, L. Palkovics, L. Gianone, Automatic detection of the lane departure of vehicles, IFAC Transportation Systems Chania, Greece (1997).

[10] D.J. LeBlanc, G.E. Johnson, P.J.T. Venhovens, G. Gerber, R. DeSonia, R. Ervin, C.F. Lin, A.G. Ulsoy, T.E. Pilutti, CAPC: a road-departure prevention system, IEEE Control Systems December (1996).

[11] C. Thorpe, M.H. Hebert, T. Kanade, S.A. Shafer, Vision and navigation for the carnegie–mellon navlab, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (3) (1988).

[12] M.A. Turk, D.G. Morgenthaler, K.D. Gremban, M. Marra, VITS—a vision system for autonomous land vehicle navigation, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (3) (1988).

[13] S.D. Buluswar, B.A. Draper, Color machine vision for autonomous vehicles, Engineering Applications of Artificial Intelligence 11 (1998) 245–256.

[14] M. Betke, E. Haritaoglu, L.S. Davis, Real-time multiple vehicle detection and tracking from a moving vehicle, Machine Vision and Applications 12 (2000) 69–83.

[15] J. Zhang, H. Nagel, Texture-based segmentation of road images, Proceedings of IEEE, Symposium on Intelligent Vehicles 94, IEEE Press, Piscataway, NJ, 1994.

[16] J. Badenas, M. Bober, F. Pla, Segmenting traffic scenes from grey level and motion information, Pattern Analysis and Applications 4 (2001) 28–38.

[17] K. Kluge, G. Johnson, Statistical characterization of the visual characteristics of painted lane markings, Proceedings of IEEE Intelligent Vehicles 95, Detroit (1995) 488–493.

[18] W. Kasprzak, An iconic classification scheme for video-based traffic sensor tasks, in: W. Skarbek (Ed.), CAIP 2001, Springer, Berlin, 2001, pp. 725–732.

[19] W. Enkelmann, G. Struck, J. Geisler, ROMA—a system for model-based analysis of road markings, Proceedings of IEEE Intelligent Vehicles 95, Detroit (1995) 356–360.

[20] A.Y. Nooralahiyan, H.R. Kirby, Vehicle classification by acoustic signature, Mathematical and Computer Modeling 27 (9–11) (1998).

[21] A. Broggi, Parallel and local feature extraction: a real-time approach to road boundary detection, IEEE Transaction on Image Processing 4 (2) (1995) 217–223.

[22] S. Beucher, M. Bilodeau, Road segmentation and obstacle detection by a fast watershed transform, Proceedings of IEEE Intelligent Vehicles 94, Paris, France October (1994) 296–301.

[23] X. Yu, S. Beucher, M. Bilodeu, Road tracking, lane segmentation and obstacle recognition by mathematical morphology, Proceedings of IEEE Intelligent Vehicles 92 (1992) 166–170.

[24] C. Kreucher, S. Lakshmanan, LANA: a lane extraction algorithm that uses frequency domain features, IEEE Transactions on Robotics and Automation 15 (2) (1999).

[25] A.L. Yuille, J.M. Coughlan, Fundamental limits of Bayesian inference: order parameters and phase transitions for road tracking, IEEE Pattern Analysis and Machine Intelligence 22 (2) (2000) 160–173. February.

[26] Y. Wang, D. Shen, E.K. Teoh, Lane detection using spline model, Pattern Recognition Letters 21 (1994) 677–689.

[27] K. Kluge, S. Lakshmanan, A deformable-template approach to lane detection, IEEE Proceedings of Intelligent Vehicles 95 (1995) 54–59.

[28] C.J. Taylor, J. Malik, J. Weber, A real time approach to stereopsis and lane-finding, IFAC Transportation Systems Chania, Greece (1997).

[29] D. Geman, B. Jedynak, An active testing model for tracking roads in satellite images, IEEE Pattern Analysis and Machine Intelligence 18 (1) (1996) 1–14.

[30] E.D. Dickmanns, Vehicle Guidance By Computer Vision, Papageorgiou Markos (Ed.), Concise Encyclopedia of Traffic and Transportation Systems.

[31] E.D. Dickmanns, B. Mysliwetz, T. Christians, An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles, IEEE Transactions on Systems, Man, and Cybernetics 20 (6) (1990).

[32] P.G. Michalopoulos, D.P. Panda, Derivation of advanced traffic parameters through video imaging, IFAC Transportation Systems Chania, Greece (1997).

[33] D.G. Morgenthaler, S.J. Hennessy, D. DeMenthon, Range-video fusion and comparison of inverse perspective algorithms in static

images, IEEE Transactions on Systems Man and Cybernetics 20 (1990) 1301–1312.
[34] E.D. Dickmanns, B.D. Mysliwetz, Recursive 3D road and relative ego-state recognition, IEEE Pattern Analysis and Machine Intelligence 14 (2) (1992) 199–213.
[35] M. Papageorgiou, Video sensors, Papageorgiou Markos (Ed.), Concise Encyclopedia of traffic and transportation systems, pp. 610–615.
[36] W. Enkelmann, Obstacle detection by evaluation of optical flow field from image sequences, Proceedings of European Conference on Computer Vision, Antibes, France 427 (1990) 134–138.
[37] B. Ross, A practical stereo vision system, Proceedings of International Conference on Computer Vision and Pattern Recognition, Seattle, WA (1993) 148–153.
[38] M. Bertozzi, A. Broggi, S. Castelluccio, A real-time oriented system for vehicle detection, Journal of System Architecture 43 (1997) 317–325.
[39] F. Thomanek, E.D. Dickmanns, D. Dickmanns, Multiple object recognition and scene interpretation for autonomous road vehicle guidance, Proceedings of IEEE Intelligent Vehicles 94, Paris, France (1994) 231–236.
[40] G.L. Foresti, V. Murino, C. Regazzoni, Vehicle recognition and tracking from road image sequences, IEEE Transactions on Vehicular Technology 48 (1) (1999) 301–317.
[41] G.L. Foresti, V. Murino, C.S. Regazzoni, G. Vernazza, A distributed approach to 3D road scene recognition, IEEE Transactions on Vehicular Technology 43 (2) (1994).
[42] H.-H. Nagel, W. Enkelmann, An investigation of smoothness constrains for the estimation of displacement vector fields from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence (1986) 565–593.
[43] T.N. Tan, G.D. Sullivan, K.D. Baker, Model-based location and recognition of road vehicles, International Journal of Computer Vision 27 (1) (1998) 5–25.
[44] D. Koller, K. Daniilidis, H. Nagel, Model-based object tracking in monocular image sequences of road traffic scenes, International Journal Computer Vision 10 (1993) 257–281.
[45] E.D. Dickmanns, V. Graefe, Dynamic monocular machine vision, Machine vision and applications 1 (1988) 223–240.
[46] Y. Park, Shape-resolving local thresholding for object detection, Pattern Recognition Letters 22 (2001) 883–890.
[47] J.M. Blosseville, C. Krafft, F. Lenoir, V. Motyka, S. Beucher, TITAN: new traffic measurements by image processing, IFAC Transportation systems, Tianjin, Proceedings (1994).
[48] Y. Won, J. Nam, B.-H. Lee, Image pattern recognition in natural environment using morphological feature extraction, in: F.J. Ferri (Ed.), SSPR&SPR 2000, Springer, Berlin, 2001, pp. 806–815.
[49] K. Shimizu, N. Shigehara, Image processing system used cameras for vehicle surveillance, IEE Second International Conference on Road Traffic Monitoring, Conference Publication Number 299 February (1989) 61–65.
[50] M. Fathy, M.Y. Siyal, An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis, Pattern Recognition Letters 16 (1995) 1321–1330.
[51] N. Hoose, Computer vision as a traffic surveillance tool, IFAC Transportation systems, Tianjin, Proceedings (1994).
[52] X. Li, Z.-Q. Liu, K.-M. Leung, Detection of vehicles from traffic scenes using fuzzy integrals, Pattern Recognition 35 (2002) 967–980.
[53] H. Moon, R. Chellapa, A. Rosenfeld, Performance analysis of a simple vehicle detection algorithm, Image and Vision Computing 20 (2002) 1–13.
[54] A. Kuehnel, Symmetry based recognition of the vehicle rears, Pattern Recognition Letters 12 (1991) 249–258. North Holland, Amsterdam.
[55] M. Fathy, M.Y. Siyal, A window-based image processing technique for quantitative and qualitative analysis of road traffic parameters, IEEE Transactions on Vehicular Technology 47 (4) (1998).
[56] D.C. Hogg, G.D. Sullivan, K.D. Baker, D.H. Mott, Recognition of vehicles in traffic scenes using geometric models, IEE, Proceedings of the International Conference on Road Traffic Data Collection, London (1984) 115–119. London.
[57] P. Klausmann, K. Kroschel, D. Willersinn, Performance prediction of vehicle detection algorithms, Pattern Recognition 32 (1999) 2063–2065.
[58] K.W. Dickinson, C.L. Wan, Road traffic monitoring using the TRIP II system, IEE Second International Conference on Road Traffic Monitoring, Conference Publication Number 299 February (1989) 56–60.
[59] G.D. Sullivan, K.D. Baker, A.D. Worrall, C.I. Attwood, P.M. Remagnino, Model-based vehicle detection and classification using orthographic approximations, Image and Vision Computing 15 (1997) 649–654.
[60] A.D. Houghton, G.S. Hobson, N.L. Seed, R.C. Tozer, Automatic vehicle recognition, IEE Second International Conference on Road Traffic Monitoring, Conference Publication Number 299 February (1989) 71–78.
[61] C.L. Wan, K.W. Dickinson, Road traffic monitoring using image processing—a survey of systems, techniques and applications, IFAC Control Computers, Communications in Transportation, Paris, France (1989).
[62] S. Mantri, D. Bullock, Analysis of feedforward-backpropagation neural networks used in vehicle detection, Transportation Research Part C 3 (3) (1995) 161–174.
[63] N. Hoose, IMPACT: an image analysis tool for motorway analysis and surveillance, Traffic Engineering Control Journal (1992) 140–147.
[64] N. Paragios, R. Deriche, Geodesic active contours and level sets for the detection and tracking of moving objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (3) (2000) 266–280.
[65] T. Aach, A. Kaup, Bayesian algorithms for adaptive change detection in image sequences using Markov random fields, Signal Processing: Image Communication 7 (1995) 147–160.
[66] N. Paragios, G. Tziritas, Adaptive detection and localization of moving objects in image sequences, Signal Processing: Image Communication 14 (1999) 277–296.
[67] J.B. Kim, H.S. Park, M.H. Park, H.J. Kim, A real-time region-based motion segmentation using adaptive thresholding and K-means clustering, in: M. Brooks, D. Corbett, M. Stumptner (Eds.), AI 2001, Springer, Berlin, 2001, pp. 213–224.
[68] M. Dubuisson, A. Jain, Contour extraction of moving objects in complex outdoor scenes, International Journal of Computer Vision 14 (1995) 83–105.
[69] N. Hoose, Computer Image Processing in Traffic Engineering, Taunton Research Studies Press, UK, 1991.
[70] J. Badenas, J.M. Sanchiz, F. Pla, Motion-based segmentation and region tracking in image sequences, Pattern Recognition 34 (2001) 661–670.
[71] C. Wohler, J.K. Anlauf, Real-time object recognition on image sequences with the adaptable time delay neural network algorithm—applications for autonomous vehicles, Image and Vision Computing 19 (2001) 593–618.
[72] K.W. Lee, S.W. Ryu, S.J. Lee, K.T. Park, Motion based object tracking with mobile, Camera Electronics Letters 34 (3) (1998) 256–258.
[73] B. Coifman, D. Beymer, P. McLauchlan, J. Malik, A real-time computer vision system for vehicle tracking and traffic surveillance, Transportation Research Part C 6 (1998) 271–288.
[74] Y.-K. Jung, Y.-S. Ho, A feature-based vehicle tracking system in congested traffic video sequences, in: H.-Y. Shum, M. Liao,

S.F. Chang (Eds.), PCM 2001, Springer, Berlin, 2001, pp. 190–197.
[75] A. Techmer, Real-time motion based vehicle segmentation in traffic lanes, in: B. Radig, S. Florczyk (Eds.), DAGM 2001, Springer, Berlin, 2001, pp. 202–207.
[76] R. Fraile, S.J. Maybank, Building 3D models of vehicles for computer vision, in: D. Huijsmans, A. Smeulders (Eds.), Visual'99, Springer, Berlin, 1999, pp. 697–702.
[77] B. Gloyer, H.K. Aghajan, K.-Y. Siu, T. Kailath, Video-based monitoring system using recursive vehicle tracking, Proceedings of IS&T/SPIE Symposium on Electronic Image: Science & Technology—Image and Video Processing (1995).
[78] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1981) 185–203.
[79] W. Enkelmann, R. Kories, H.-H. Nagel, G. Zimmermann, An experimental investigation of estimation approaches for optical flow fields, in: W.N. Martin, J.K. Aggarwal (Eds.), Motion Understanding: Robot and Human Vision, Kluwer, Dordrecht, 1987, pp. 189–226.
[80] R. Kories, H.-H. Nagel, G. Zimmermann, Motion detection in image sequences: an evaluation of feature detectors, Proceedings of International Joint Conference on Pattern Recognition, Montreal (1984) 778–780.
[81] W. Enkelmann, Interpretation of traffic scenes by evaluation of optical flow fields from image sequences, IFAC Control Computers, Communications in Transportation, Paris, France (1989).
[82] H.H. Nagel, Constraints for the estimation of displacement vector fields from image sequences, Proceedings of Intelligent Joint Conference on Artificial Intelligence, Karlsruhe/FRG August (1983) 945–951.
[83] A. Giachetti, M. Campani, V. Torre, The use of optical flow for road navigation, IEEE Transactions on Robotics and Automation 14 (1) (1998).
[84] R. Kories, G. Zimmermann, Workshop on Motion: Representation and Analysis, Kiawah Island Resort, Charleston/SC, Workshop on Motion: Representation and Analysis, Kiawah Island Resort, Charleston/SC, May, IEEE Computer Society Press, 1986, pp. 101–106.
[85] S.A. Velastin, J.H. Yin, M.A. Vicencio-Silva, A.C. Davies, R.E. Allsop, A. Penn, Image processing for on-line analysis of crowds in public areas, IFAC Transportation systems, Tianjin, Proceedings (1994).
[86] W. Enkelmann, Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences, Computer Vision, Graphics, and Image Processing (1988) 150–177.
[87] S. Carlsson, J.O. Eklundh, Object detection using model-based prediction and motion parallax, Proceedings of Europe Conference on Computer Vision, Antibes, France 427 (1990) 297–306.
[88] A.R. Bruss, B.K.P. Horn, Passive navigation, Computer Vision, Graphics, and Image Processing 21 (1983) 3–20.
[89] L. Dreschler, H.-H. Nagel, Volumetric model and 3D-trajectory of a moving car derived from monocular TV—frame sequences of a street scene, Computer Vision, Graphics, and Image Processing 20 (1982) 199–228.
[90] M. Irani, P. Anandan, A unified approach to moving object detection in 2D and 3D scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (6) (1998) 577–589.
[91] M. Irani, B. Rousso, S. Peleg, Recovery of egomotion using region alignment, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (3) (1997) 268–272.
[92] A. Nagai, Y. Kuno, Y. Shirai, Detection of moving objects against a changing background, Systems and Computer in Japan 30 (11) (1999) 107–116.
[93] S.M. Smith, J.M. Brady, A scene segmenter; visual tracking of moving vehicles, Engineering Applications of Artificial Intelligence 7 (2) (1994) 191–204.
[94] H.A. Mallot, H.H. Bulthoff, J.J. Little, S. Bohrer, Inverse perspective mapping simplifies optical flow computation and obstacle detection, Biological Cybernetics 64 (1991) 177–185.
[95] S. Tsugawa, Vision-based vehicles in Japan: machine vision systems and driving control systems, IEEE Transactions on Industrial Electronics 41 (4) (1994) 398–405.
[96] D. Pomerleau, T. Jocchem, Rapidly adaptive machine vision for automated vehicle steering, IEEE Machine Vision April (1996) 19–27.
[97] L. Matthies, Stereo vision for planetary rovers: stochastic modeling to near real-time implementation, International Journal of Computer Vision 8 (1992) 71–91.
[98] J. Schick, E.D. Dickmanns, Simultaneous estimation of 3D shape and motion of objects by computer vision, Proceedings of IEEE Workshop on Visual Motion, Princeton, NJ October (1991) 256–261.
[99] S. Mammar, J.M. Blosseville, Traffic variables recovery, IFAC Transportation Systems, Tianjin, Proceedings (1994).
[100] P.G. Michalopoulos, R.D. Jacobson, C.A. Anderson, T.B. DeBruycker, Automatic Incident Detection Through Video Image Processing, Traffic Engineering + Control February (1993) 66–75.
[101] P.G. Michalopoulos, Vehicle detection video through image processing: the autoscope system, IEEE Transactions on Vehicular Technology 40 (1) (1991).
[102] J.G. Postaire, P. Stelmaszyk, P. Bonnet, A visual surveillance system for traffic collision avoidance control, IFAC Transportation Symposium International Federation of Automatic Control, Laxenburg, Austria (1986).
[103] S. Takaba, et al., A traffic flow measuring system using a solid state sensor, Proceedings of IEE Conference on Road Traffic Data Collection, London, UK (1984).
[104] A. Rourke, M.G.H. Bell, Applications of low cost image processing technology in transport, Proceedings of the World Conference on Transport Research, Japan (1989) 169–183.
[105] Z. Zhu, G. Xu, B. Yang, D. Shi, X. Lin, VISATRAM: a real-time vision system for automatic traffic monitoring, Image and Vision Computing 18 (2000) 781–794.
[106] K. Kluge, Extracting road curvature and orientation from image edge points without perceptual grouping into features, Proceedings of IEEE Intelligent Vehicles Symposium'94 (1994) 109–111.
[107] S.K. Kenue, LANELOK: an algorithm for extending the lane sensing operation range to 100 feet, Procedings of SPIE-Mobile Robots V 1388 (1991) 222–233.
[108] M. Beauvais, S. Lakshmanan, CLARK: a heterogeneous sensor fusion method for finding lanes and obstacles, Image and Vision Computing 18 (5) (2000) 397–413.
[109] D. Pomerleau, Ralph: rapidly adapting lateral position handler, Proceedings of IEEE Intelligent Vehicles'95, Detroit, MI (1995) 506–511.
[110] Y. Goto, K. Matsuzaki, I. Kweon, T. Obatake, CMU sidewalk navigation system: a blackboard outdoor navigation system using sensor fusion with color-range images, Proceedings of First Joint Conference ACM/IEEE November (1986).
[111] J.D. Crisman, C.E. Thorpe, SCARF: a color vision system that tracks roads and intersections, IEEE Transactions on Robotics and Automation 9 (1) (1993).
[112] T.M. Jochem, D.A. Pomereau, C.E. Thorpe, Maniac, a next generation neurally based autonomous road follower, Proceedings of International Conference on Intelligent Autonomous Systems: IAS-3, Pittsburgh, PA February (1993).
[113] H.H. Nagel, F. Heimes, K. Fleischer, M. Haag, H. Leuck, S. Noltemeier, Quantitative comparison between trajectory estimates obtained from a binocular camera setup within a moving road vehicle and from the outside by a stationary monocular camera, Image and Vision Computing 18 (2000) 435–444.

[114] Q. Zheng, R. Chellappa, Automatic feature point extraction and tracking in image sequences for arbitrary camera motion, International Journal of Computer Vision 15 (1995) 31–76.

[115] D. Coombs, M. Herman, T. Hong, M. Nashman, Real-time obstacle avoidance using central flow divergence, and peripheral flow, IEEE Transactions on Robotics and Automation 14 (1) (1998) 49–59.

[116] N. Kehtarnavaz, N.C. Griswold, J.S. Lee, Visual control of an autonomous vehicle (BART)—the vehicle-following problem, IEEE Transactions on Vehicular Technology 40 (3) (1991) 654–662.

[117] S.E. Shladover, D.A. Desoer, J.K. Hedrick, M. Tomizuka, J. Walrand, W.-B. Zhang, D.H. McMahon, H. Peng, S. Sheikholeslam, N. McKeown, Automatic vehicle control developments in the PATH program, IEEE Transactions on Vehicular Technology 40 (1) (1991) 114–129.

[118] S. Kato, S. Tsugawa, Cooperative driving of autonomous vehicles based on localization, inter-vehicle communications and vision systems, JSAE Review 22 (2001) 503–509.

[119] M.M. Trivedi, M.A. Abidi, R.O. Eason, R.C. Gonzalez, Developing robotics systems with multiple sensors, IEEE Transactions on Systems, Man, and Cybernetics 20 (6) (1990) 1285–1300.

[120] F. Chavand, E. Colle, Y. Chekhan, E.C. N'zi, 3D measurements using a video and a range finder, IEEE Transactions on Instrumentation and Measurement 46 (6) (1997) 1229–1235.

[121] M.J. Magee, B.A. Boyter, C.H. Chien, J.K. Aggarwal, Experiments in intensity guided range sensing recognition of three-dimensional objects, IEEE Pattern Analysis and Machine Intelligence 7 (1985) 629–637.

[122] P.J. Burt, Smart sensing within a pyramid vision machine, Proceedings of IEEE 76 (1988) 1006–1115.

[123] A. Jain, T. Newman, M. Goulish, Range-intensity histogram for segmenting LADAR images, Pattern Recognition Letters 13 (1992) 41–56.

[124] D.A. Pomerleau, Advances in Neural Information Processing, Advances in Neural Information Processing, 1, Morgan-Kaufman, San Francisco, CA, 1989, pp. 305–313.

[125] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, W. Seelen, An image processing systems for driver assistance, Image and Vision Computing (18) (2000) 367–376.

[126] H. Schodel, Utilization of fuzzy techniques in intelligent sensors, Fuzzy Sets and Systems 63 (1994) 271–292.

[127] F. Sandakly, G. Giraudon, 3D scene interpretation for a mobile robot, Robotics and Autonomous Systems 21 (1997) 399–414.

[128] Y.-G. Wu, J.-Y. Yang, K. Liu, Obstacle detection and environment modeling based on multisensor fusion for robot navigation, Artificial Intelligence in Engineering 10 (1996) 323–333.