Artificial Vision in Road Vehicles

MASSIMO BERTOZZI, ASSOCIATE MEMBER, IEEE, ALBERTO BROGGI, ASSOCIATE MEMBER, IEEE, MASSIMO CELLARIO, ALESSANDRA FASCIOLI, MEMBER, IEEE, PAOLO LOMBARDI, AND MARCO PORTA

Invited Paper

The last few decades witnessed the birth and growth of a new sensibility to transportation efficiency. In particular, the need for efficient and improved people and goods mobility pushed researchers to address the problem of intelligent transportation systems. This paper surveys the most advanced approaches to the (partial) customization of road following task, using on-board systems based on artificial vision. The functionalities of lane detection, obstacle detection and pedestrian detection are described and classified, and their possible application on future road vehicles is discussed.

Keywords—*Automatic vehicles guidance, intelligent transportation systems, intelligent vehicles, machine vision.*

I. INTRODUCTION

Problems concerning traffic mobility, safety, and energy consumption have become more serious in most developed countries in recent years. The endeavors to solve these problems have triggered the interest toward new fields of research and application, such as automatic vehicle driving, in which new techniques are investigated for the entire or partial automation of driving tasks. A recently defined comprehensive and integrated system approach, referred to as intelligent transportation systems (ITS), links the vehicle, the infrastructure, and the driver to make it possible to achieve more mobile and safer traffic conditions by using state-of-the-art electronic communication and computer-controlled technology.

Over time, the ITS research community expects that intelligent vehicles will advance in three primary ways: in the capabilities of in-vehicle systems, in the sophistication of the driver–vehicle interface, and in the ability of vehicles to communicate with each other and a smart infrastructure [1].

Smart vehicles will be able to give route directions, sense objects, warn drivers of impending collisions, automatically signal for help in emergencies, keep drivers alert, and may ultimately be able to take over driving.

In fact, ITS technologies may provide vehicles with different types and levels of "intelligence" to complement the driver. Information systems expand the driver's knowledge of routes and locations. Warning systems, such as collision-avoidance technologies, enhance the driver's ability to sense the surrounding environment. Driver assistance and automation technologies simulate the driver's sensor-motor system to operate a vehicle temporarily during emergencies or for prolonged periods.

The timing of "human-centered" intelligent vehicles' arrival on the market, however, will depend on the resolution of technical and cost constraints for some advanced concepts, such as collision-avoidance and automated systems, manufacturers' interest, production lead-times, and consumer demand.

Human-centered intelligent vehicles hold a major potential for industry. Since 1980, major car manufacturers and other firms have been developing computer-based in-vehicle navigation systems. Today, most developed/developing systems around the world have included more complex functions to help people drive their vehicles safely and efficiently.

New information and control technologies that make vehicles smarter are now arriving on the market either as optional equipment or as specialty after-market components. These technologies are being developed and marketed to increase driver safety, performance, and convenience. However, these disparate individual components have yet to be integrated to create a coherent intelligent vehicle that complements the human driver, fully considering his requirements, capabilities and limitations.

A fully intelligent vehicle must work cooperatively with the driver [1]: an intelligent system senses its environment and acts to reach its objectives: its interaction-communication channels have a big influence on the type of intelligence it can display [2].

New uncoordinated technologies could deliver excessive, competing, or contradictory messages and demands that might distract, confuse and overwhelm the driver, overloading his limited cognitive resources and eventually leading to a decrease in his own performance and safety.

Clearly there is a need in the research community for quantitative and objective performance metrics to define and structure this problem domain, for describing and evaluating products in future competitive markets when research is reduced to technology and commercialized.

In the last two decades, government institutions have activated initial explorative phases by means of various projects worldwide, involving a large number of research units who worked cooperatively, producing several prototypes and solutions, based on rather different approaches.

In Europe, the PROMETHEUS project (PROgraM for a European Traffic with highest Efficiency and Unprecedented Safety) started this explorative stage in 1986. The project involved more than 13 vehicle manufacturers and several research units from governments and universities of 19 European countries. Within this framework, a number of different ITS approaches were conceived, implemented, and demonstrated.

In the United States, a great deal of initiatives were launched to address the mobility problem, involving universities, research centers, and automobile companies. After this pilot phase, in 1995 the US government established the National Automated Highway System Consortium (NAHSC) [3], and launched the Intelligent Vehicle Initiative (IVI) right after in 1997.

In Japan, where the mobility problem is even more intense and evident, some vehicle prototypes were also developed within the framework of different projects. Similarly to the US case, in 1996 the Advanced Cruise-Assist Highway System Research Association (AHSRA) was established amongst a large number of automobile industries and research centers [4], which developed different approaches to the problem of automatic vehicle guidance.

As a whole, the main results of this first stage provided a deep analysis of the problem and the development of a feasibility study to understand the requirements and possible effects of ITS technology applications.

The ITS field is now entering its second phase characterized by a maturity in approaches and by new technological possibilities which allow the development of the first experimental products. A number of prototypes of intelligent vehicles have been designed, implemented, and tested on the road. The design of these prototypes has been preceded by the analysis of solutions deriving from similar and close fields of research, and has produced a great flourishing of new ideas, innovative approaches, and novel *ad hoc* solutions. Robotics, artificial intelligence, computer science, computer architectures, telecommunications, control and automation, and signal processing are just some of the principal research areas from which the main ideas and solutions were first derived. Initially, underlying technological devices—such as head-up displays, infrared cameras, radars, and sonars—derived from expensive military applications, but, thanks to the increased interest in these applications and

to the progress in industrial production, today's technology offers sensors, processing systems, and output devices at very competitive prices. In order to test a wide spectrum of different approaches, these automatic vehicles prototypes are equipped with a large number of different sensors and computing engines.

Section II of this paper describes the motivations which underlie the development of vision-based intelligent vehicles and illustrates their requirements and peculiarities. Section III surveys the most common approaches to Road Following developed worldwide. Section IV briefly introduces a few significant architectural issues, while Section V outlines our perspectives in the evolution of intelligent vehicles.

II. IMPROVING ON-ROAD MOBILITY

A. Intelligent Vehicles Versus Intelligent Infrastructures

The enhancement of future vehicles' efficiency can be achieved acting both on infrastructures and on vehicles. Depending on the specific application, either choice possesses advantages and drawbacks. Enhancing road infrastructure may yield benefits to transportation architectures based on repetitive and prescheduled routes, such as public transportation and industrial robotics. On the other hand, extended road networks for private vehicles would require a complex and extensive organization and maintenance which may become cumbersome and extremely expensive: an *ad hoc* structuring of the environment can only be considered for a reduced subset of the road network, for example, a fully automated highway on which only automatic vehicles—public or private—can drive.

A great deal of research has been focused on advancement in vehicle safety and automation systems, applied to different major classes of vehicles: light, commercial, transit, and specialty vehicles; particular attention has been dedicated to selected problem areas, as "prime candidates" for improving vehicle safety and efficiency: lane keeping and road departure warning, collision avoidance (intersection, merge, rear-end, vehicles, obstacles, pedestrians), vision enhancement, vehicle stability, driver condition monitoring, and safety-impacting in-vehicle technology integration.

In this paper, only in-vehicle control and automation applications are considered, while road infrastructure, inter-vehicle communication, satellite communication, information systems, and driver-vehicle interface issues are not covered.

Any on-board system for ITS applications needs to meet some important requirements.

- The final system, installed on a commercial vehicle, must be sufficiently robust to adapt to different conditions and changes of environment, road, traffic, illumination, and weather. Moreover, the hardware system needs to be resistant to mechanical and thermal stress.
- On-board systems for ITS applications are safety-critical and require a high degree of reliability: the project has to be thorough and rigorous during all its phases, from requirements specification to design and implementation. An extensive phase of testing and validation is therefore of paramount importance.
- For marketing reasons, the design of an ITS system is driven by strict cost criteria (it should cost no more than

10% of the vehicle price), thus requiring a specific engineering phase. Operative costs (such as power consumption) need to be kept low as well, since vehicle performance should not be affected by the use of ITS apparata.

- The system's hardware and sensors have to be kept compact in size and should not disturb car styling.
- The design of the driver–vehicle interface (the place where the driver interacts physically and cognitively with the vehicle) is critical. When giving drivers access to ITS systems inside the vehicle, designers must not only consider safety (i.e., overloading the driver's information-processing resources), but also usability and driver acceptance [5]: interfaces will need to be intelligent and user-friendly, effective, and transparent to use; in particular, a full understanding of the subtle tradeoffs of multimodal interface integration will require significant research [2].

B. Active Versus Passive Sensors

Among the sensors widely used in indoor robotics, tactile sensors and acoustic sensors are of no use in automotive applications because of vehicles' speed and reduced detection range.

Laser-based sensors and millimeter-wave radars detect the distance of objects by measuring the travel time of a signal emitted by the sensors themselves and reflected by the object, and are therefore classified as *active sensors*. Their main common drawbacks consist of low spatial resolution and slow scanning speed. However, millimeter-wave radars are more robust to rain and fog than laser-based radars, though more expensive.

Vision-based sensors are defined as *passive sensors* and have an intrinsic advantage over laser and radar sensors: the possibility of acquiring data in a noninvasive way, thus not altering the environment (image scansion is performed fast enough for ITS applications). Moreover, they can be used for some specific applications for which visual information plays a basic role (such as lane markings localization, traffic signs recognition, and obstacle identification) without requiring any modifications to road infrastructures. Unfortunately, vision sensors are less robust than millimeter-wave radars in foggy, night, or direct sunshine conditions.

Active sensors possess some specific peculiarities which result in advantages over vision-based sensors, in this specific application: they can measure some quantities, such as movement, in a more direct way than vision and require less performing computing resources, as they acquire a considerably lower amount of data.

Nevertheless, besides the problem of environment pollution, the wide variation in reflection ratios caused by different reasons (such as obstacles shape or material) and the need for the maximum signal level to comply with some safety rules, the main problem in using active sensors is represented by interference among sensors of the same type, which could be critical for a large number of vehicles moving simultaneously in the same environment, as, example, in the case of autonomous vehicles traveling on intelligent highways.

Hence, foreseeing a massive and widespread use of autonomous sensing agents, the use of passive sensors, such as cameras, obtains key advantages over the use of active ones.

Obviously machine vision does not extend sensing capabilities besides human possibilities in very critical conditions (e.g., in foggy weather or at night with no specific illumination), but can, however, help the driver in case of failure, for example, due to a lack of concentration or due to drowsiness.

C. Vision-Based Intelligent Vehicles

Some important issues must be carefully considered in the design of a vision system for automotive applications. In the first place, ITS systems require faster processing than other applications, since vehicle speed is bounded by the processing rate. The main problem that has to be faced when real-time imaging is concerned and which is intrinsic to the processing of images is the large amount of data—and therefore computation—involved. As a result, specific computer architectures and processing techniques must be devised in order to achieve real-time performance. Nevertheless, since the success of ITS apparata is tightly related to their cost, the computing engines cannot be based on expensive processors. Therefore, either of-the-shelf components or *ad hoc* dedicated low-cost solutions must be considered.

Secondly, in the automotive field, no assumptions can be made on key parameters, for example, scene illumination or contrast, which are directly measured by the vision sensor. Hence, the subsequent processing must be robust enough to adapt to different environmental conditions (such as sun, rain, or fog) and to their dynamic changes (such as transitions between sun and shadow, or the entrance or exit from a tunnel).

Furthermore, other key issues, such as the robustness to vehicle's movements and drifts in the camera's calibration, must be handled as well.

However, recent advances in both computer and sensor technologies promote the use of machine vision also in the intelligent vehicles field. The developments in computational hardware, such as a higher degree of integration and a reduction of the power supply voltage, permit the production of machines that can deliver a high computing power, with fast networking facilities, at an affordable price. Current technology allows the use of SIMD-like processing paradigms even in general-purpose processors, such as the new generation of processors that include multimedia extensions.

In addition, current cameras include new important features that permit the solution of some basic problems directly at sensor level. For example, image stabilization can be performed during acquisition, while the extension of camera dynamics allows one to avoid the processing required to adapt the acquisition parameters to specific light conditions. The resolution of the sensors has been drastically enhanced, and, in order to decrease the acquisition and transfer time, new technological solutions can be found in CMOS sensors, such as the possibility of dealing with pixels independently as in traditional memories. Another key advantage of CMOS-based sensors is that their integration on the processing chip seems to be straightforward.

Many different parameters must be evaluated for the design and choice of an image acquisition device. First of all,

some parameters tightly coupled with the algorithms regard the choice of monocular versus binocular (stereo) vision and the sensors' angle of view (some systems adopt a multi-camera approach by using more than one camera with different viewing angles, e.g., fish eye or zoom). The resolution and the depth (number of bit/pixel) of the images have to be selected as well (this also includes the selection of color versus monochrome images).

Other parameters—intrinsic to the sensor—must be considered. Although the frame rate is generally fixed for CCD-based devices (25 or 30 Hz), the dynamics of the sensor is of basic importance: conventional cameras allow an intensity contrast of 500:1 within the same image frame, while most ITS applications require a 10 000:1 dynamic range for each frame and 100 000:1 for a short image sequence. Different approaches have been studied to meet this requirement, ranging from the use of CMOS-based cameras with a logarithmically compressed dynamic [6], [7] to the interpolation and superimposition regarding values of two subsequent images taken from the same camera [8].

In conclusion, although extremely complex and highly demanding, computer vision is a powerful means for sensing the environment and has been widely employed to deal with a large number of tasks in the automotive field, thanks to the great deal of information it can deliver (it has been estimated that humans perceive visually about 90% of the environment information required for driving).

III. THE ROAD FOLLOWING DRIVING TASK

Among the complex and challenging tasks of future road vehicles is road following. It is based on: lane detection (which includes the localization of the road, the determination of the relative position between vehicle and road, and the analysis of the vehicle's heading direction), and obstacle detection (which is mainly based on localizing possible obstacles on the vehicle's path). Moreover, growing demand is also posed on the problem of pedestrian detection, since safety and avoidance of car crushes with pedestrians is a central concern for future systems.

In this section, a survey on the most common approaches to lane detection, obstacle detection and pedestrian detection is presented, focusing on vision-based systems.

A. Lane Detection

In most prototypes of autonomous vehicles developed worldwide, lane following is divided into the following two steps: initially the relative position of the vehicle with respect to the lane is computed and then actuators are driven to keep the vehicle in the correct position.

Conversely, some early systems were not based on the preliminary detection of the road's position, but obtained the commands to be issued to the actuators (steering wheel angles) directly from visual patterns detected in the incoming images. For example, the Autonomous Land Vehicle In a Neural Net (ALVINN) system is based on a neural net approach: it is able to follow the road after a training phase with a large set of images [9].

Nevertheless, since the knowledge of the lane position can be conveniently exploited by other driving assistance functions, the localization of the lane is generally performed.

A few systems have been designed to handle completely unstructured roads: for example, the Supervised Classification Applied to Road Following (SCARF) [10] and POSTECH Road Vehicle (PVR III) [11] systems are based on the use of color cameras and exploit the assumption of a homogeneously colored road to extract the road region from the images.

More generally, however, lane detection has been reduced to the localization of specific features such as markings painted on the road surface. This restriction eases the detection of the road, nevertheless two basic problems must be faced.

- The presence of shadows (projected by trees, buildings, bridges, or other vehicles) produces artifacts onto the road surface and thus alterates the road texture. Most research groups face this problem by using highly sophisticated image filtering algorithms. When lane markings are not well visible (because of low contrast, shadows, bad weather conditions, etc.), the use of pattern-based techniques can be helpful. The system developed at the Toyota Central R&D Labs, for example, is based on a voting method, in which lane marking patterns are generated and provided. After ordinary edge extraction, edge points are matched to each pattern. At the end of the process, the patterns with the greater number of votes are chosen as the best approximations of the left and right lane markings [12]. Other algorithms exploit the processing of color images, this is the case of the Michigan Offroad Sensor Fusing Experimental Testbed (MOSFET) autonomous vehicle which uses a color segmentation algorithm that maximizes the contrast between lane markings and the road [13].
- Other vehicles on the path partly occlude the visibility of the road and therefore of road markings as well.

To cope with this problem, some systems have been designed to investigate only a small portion of the road ahead of the vehicle where the absence of other vehicles can be assumed. As an example, the LAKE and SAVE autonomous vehicles rely on the processing of the image portion corresponding to the nearest 12 m of road ahead of the vehicle, and it has been demonstrated that this approach is able to safely maneuver the vehicle on highways and even on belt ways or ramps with a bending radius down to 50 m [14].

On the other hand, some systems solve the occlusion problem by combining lane detection with obstacle detection. For example, the Rapidly Adapting Lateral Position Handler (RALPH) system reduces the portion of the image to be processed according to the result of a radar-based obstacle detection module [15].

The algorithm developed by General Dynamics Robotic Systems and The Ohio State University exploits the road gray-level histogram to detect lane markings, which are then analyzed using a decision tree. A histogram-based segmenta-

tion classifies the objects in the scene as road, lane markings candidates or obstacle (vehicle) candidates [16].

In other cases, the search area for lane markings detection is determined first. The research group of the Laboratoire Central des Ponts-et-Chaussees de Strasbourg assumes that there should always be a chromatic contrast between road and off-road areas (or obstacles), at least in one color component; the concept of chromatic saturation is used to separate the components [17].

Since lane detection is generally based on the localization of *specific patterns* (lane markings), it can be performed with the analysis of a single still image. In addition, some assumptions may help and/or speed up the detection process.

- Due to both physical and continuity constraints, the processing of the whole image can be replaced by the analysis of specific regions of interest only (the so-called focus of attention), in which the features of interest are more likely to be found. This is a generally followed strategy that can be adopted using the results of previously processed frames or assuming an *a priori* knowledge on the road environment.

In some approaches, in particular, windows of interest (WOIs) are determined dynamically by means of statistical methods. For example, the system developed at LASMEA selects the proper window according to the current state and previously detected WOIs [18].

A system developed by the Robert Bosch GmbH research group, on the other hand, employs a model both for the road and the vehicle's dynamic to determine the road portion where it is most likely to find lane markings [19].

- The assumption of a fixed or smoothly varying lane width allows the enhancement of the search criterion, limiting the search to almost parallel lane markings.

As an example, on the PVR III vehicle, lane markings can be detected using both neural networks and simple vision algorithms: two parallel stripes of the acquired image are selected and filtered using Gaussian masks and zero crossing to find vertical edges. The result is matched against a given model (a typical road pattern with parallel lane markings) to compute a steering angle and a fitness evaluation indicating the confidence in the result [11].

Analogously, the RALPH system is based on the processing of the image portion corresponding to the road about 20–70 m ahead of the vehicle, depending on the vehicle's speed and obstacles presence. The perspective effect is removed from this portion of the image and the determination of the curvature is carried out according to a number of possible curvature models for a specific road template featuring parallel road markings [15].

- The reconstruction of road geometry can be simplified by assumptions on its shape.

The research groups of the Universität der Bundeswehr [20], Daimler-Benz [21] and Robert Bosch GmbH [22] base their road detection functionality on a specific road model: lane markings are modeled as clothoids. In a clothoid the curvature depends linearly on the curvilinear reference. This model has the advantage that the knowledge of two parameters only allows the full localization of lane markings and the computation of other parameters like the lateral offset within the lane, the lateral speed with respect to the lane, and the steering angle. Another system based on a clothoidal modelization of lane markings is the one developed at The Ohio State University, where a dynamic programming optimization method is used to chose among center-line candidates representing the actual geometry of the road [23].

Other research groups use a polynomial representation for lane markings. In the MOSFET autonomous vehicle, for instance, lane markings are modeled as parabolas [13]. A simplified Hough transform is used to accomplish the fitting procedure.

Similarly, a preliminary version of the lane detection system developed at The Ohio State University Center for Intelligent Transportation Research relies on a polynomial curve [24]. It assumes a flat road with either continuous or dashed bright lane markings. The history of previously located lane markings is used to determine the region of interest, thus reducing the portion of the image to be processed. The algorithm extracts the significant bright regions from the image plane and stores them in a vector list. Qualitative parameters such as lines convergence at infinity or lane width, known or estimated, are used to extract the candidate lane markings from the list. Finally, in order to handle also dashed lines, a low-order polynomial curve is fitted to the computed vectors.

As another example, a system developed at The Université Blaise Pascal exploits a polynomial road modelization to calculate the impact distance from the vehicle to the nearest road side. Such distance is obtained by considering the intersection between the straight line trajectory followed in case of a driver loss of control and the polynomial function describing the road side [25].

Recently, a research group from the University of Michigan has proposed to use concentric circles to represent lane boundaries. Since, at least in the United States., lanes are actually laid on concentric circles, circular shape models can in fact be better choices than polynomial approximations [26].

On the contrary, other systems adopt a more generic model for the road. The ROMA vision-based system uses a contour-based method [27]. A dynamic road model permits the processing of small portions of the acquired images therefore enabling real-time performance. Actually, only straight or small curved roads without intersections are included in this model. Images are processed using a gradient-based filter and a programmable threshold. The road model is used to follow contours formed by pixels that feature a significant gradient direction value.

The CyCab electric vehicle uses an edge linking process based on Contour Chains and Causal Neighborhood Windows (areas of interest connected to edge elements). After an initial segmentation phase, the longest chains with slope angles close to 45 and 135 degrees are searched for, as they represent the most probable candidates for left and right lanes [28].

Table 1
Pros and Cons of the Most Typical Assumptions in Lane Detection

	Pros	Cons
Focus of attention	Analysis of a small image portion, fast processing, real-time performance, low-cost hardware	Does not fit large features, choice of the region of interest is critical
Fixed lane width	Enhancement of the search criterion (parallel line markings), robustness to shadows and occlusions	Does not match roads with variable lane width
Road shape	Strong model robust w.r.t. shadows and occlusions, eases road geometry reconstruction, simplifies vehicle control	Requires fitting with complex equations, high computational power needed, fails if the road does not match the model
A priori knowledge on the road surface/slope	Simplifies mapping between image pixels and world coordinates (e.g., determination of obstacle distance, road curvature)	Hypotheses (e.g., flat road) are not always met in real cases, approximations of recalibration needed

Similarly, the system implemented at the Transportation College of Jilin University of Technology is based on a linear lane model, where road markers are reconstructed as sequences of straight lines [29].

A generic triangular road model was originally tested on the MOB-LAB experimental vehicle by the research groups of the University of Parma [30] and the Istituto Elettrotecnico Nazionale "G. Ferraris," CNR, Italy [31].

- The knowledge of the specific camera calibration together with the assumption of an *a priori* knowledge on the road (i.e., a flat road without bumps) can be exploited to ease the localization of features and/or to simplify the mapping between image pixels and their correspondent world coordinates.

The majority of previously discussed systems exploit the assumption of a flat road in front of the vehicle in the determination of obstacle distance or road curvature, once the specific features of interest have been localized in the acquired image. The Generic Obstacle and Lane Detection (GOLD) system [32] implemented on the ARGO autonomous vehicle and the already mentioned RALPH system, however, exploit this assumption also in the lane determination process. In fact, in both cases the lane markings detection is performed in a different image domain, representing a bird's eye view of the road, which can be obtained thank to the flat road assumption.

Table 1 summarizes the pros and cons of the assumptions on which the most common approaches to lane detection rely.

Other methods, based on statistical approaches, have been experimented to cope with unfriendly lighting and weather conditions. In the already quoted system implemented at LASMEA, for instance, the search for road markers is carried out as an iterative process, where continuous updates of the lane model and the size of areas of interest allow the

lane detection task to be relatively noise insensitive [33], [18].

When the processing is aimed not only at a mere lane detection but also at lane markings tracking, the temporal correlation between consecutive frames can be used either to ease the feature determination or to validate the result of the processing. The lane detection module implemented and tested on the ARGO vehicle falls in the first case as it restricts the image portion to be analyzed to the nearest neighborhood of the markings previously detected [34]. In a different manner, the lane detection module developed by the research group at the Istituto Elettrotecnico Nazionale "G. Ferraris" uses the result of previous computations to validate the current one, once lane markings are found by means of a triangular model [31].

B. Obstacle Detection

The criteria used for the detection of obstacles depend on the definition of what an obstacle is (see Fig. 1).

In some systems, determining obstacles is limited to the localization of vehicles, which is then based on a search for specific patterns, possibly supported by other features, such as shape, symmetry, or the use of a bounding box.

Conversely, the Obstacle Detection algorithm developed at the Universität der Bundeswehr is based both on an edge detection process and on obstacle modelization; the system is able to detect and track up to twelve objects around the vehicle. The continuously updated obstacle variables are: distance, direction, relative speed, relative acceleration, lateral position, lateral speed, and size [20].

Analogously, in the vehicle detecting and tracking system developed by the R&D Group at NEC Corporation, the search for possible cars and trucks is composed of two stages. After an edge-based potential vehicle identification procedure, a vehicle validation process is carried out. By
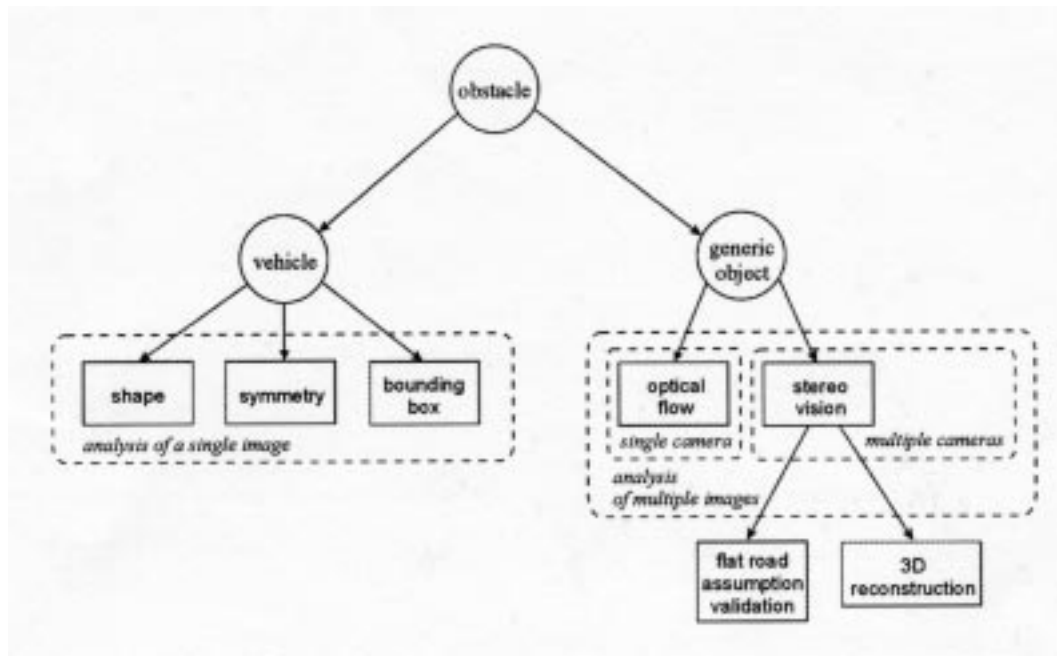
Fig. 1. Depending on the definition of obstacle, different techniques are used.

exploiting characteristics such as symmetry, shadow underneath the vehicle, and differences in gray-level average intensities, false detections can be usually removed, also in case of bad weather conditions [35].

When Obstacle Detection is limited to the localization of specific patterns, as in the previous examples, processing can be based on the analysis of a single still image, in which relevant features are searched for. For example, in another system developed at the Istituto "G. Ferraris," a strategy is proposed which formulates obstacle hypotheses by region segmentation algorithms. The hypotheses are then validated by matching edge segmentations of these regions with a dynamic model of the vehicle, which takes into account the various rear parts of a typical car (rear window, bumper, license plate, etc.) [36].

However, unfortunately, the pattern-based approach is not successful when an obstacle does not match the model.

A more general definition of obstacle, which obviously leads to more complex algorithmic solutions, identifies as an obstacle any object that obstructs the vehicle's driving path or, in other words, anything rising out significantly from the road surface. In this case, Obstacle Detection is reduced to identifying the free-space (the area in which the vehicle can safely move) instead of recognizing specific patterns.

Due to the general applicability of this definition, the problem is dealt with using more complex techniques; the most common ones are based on the processing of two or more images, such as:

- the analysis of optical flow field;
- the processing of nonmonocular images.

The optical flow-based technique requires the analysis of a sequence of two or more images: a two-dimensional (2-D) vector is computed in the image domain, encoding the horizontal and vertical components of the velocity of each pixel.

The result can be used to compute ego-motion, which in some systems is directly extracted from odometry; obstacles can be detected by analyzing the difference between the expected and real velocity fields.

As an example, the ROMA system integrates an obstacle detection module that is based on the use of an optical flow technique in conjunction with data coming from an odometer [37].

Similarly, the A Scene Segmenter Establishing Tracking v2 (ASSET-2) is a complete real-time vision system for segmentation and tracking of independently moving objects. Its main feature is that it does not require any camera calibration. It tracks down objects and is capable of correctly handling occlusions amongst obstacles and automatically tracks down each new object that enters the scene. ASSET-2 initially builds a sparse image flow field and then segments it into clusters that feature homogeneous flow variation. Temporal correlation is used to filter the result, therefore improving accuracy [38].

As another example, the system developed at Kumamoto University uses a technique based on Focus of Expansion (FOE) to carry out camera three-dimensional (3-D) motion analysis. By removing background changes, moving objects can be detected and tracked in the scene. The tracking method is based on a continuous estimation of moving objects' "vitality" and "reliability" values and can deal with up to more than 30 objects in real time. Vitality of a tracked object increases when there is a sequence of template matching successes, while decreases (eventually reaching zero) after a sequence of bad matchings, indicating that the object cannot be identified further. Reliability, instead, is the quality of a template matching [39].

Still exploiting monocular vision, other research groups base their techniques on simpler principles. For example, at LASMEA, a system has been developed which regulates the

Table 2
Comparison of the Different Approaches to Obstacle Detection

	Pros	Cons
Analysis of a single image	Simple algorithmic solutions, fast processing, does not suffer from vehicle movement	Loss of info about scene's depth unless specific assumptions are made, not successful when obstacles do not match the model
Optical flow	Detection of generic objects, allows computation of ego-motion and obstacles' relative speed	Computationally complex, sensitive to vehicle movements and drifts in calibration, fails when both vehicle and obstacle have small or null speed
Stereo vision	Detection of generic objects, allows 3D reconstruction	Computationally complex (still domain specific constraints may reduce complexity), sensitive to vehicle movements and drifts in calibration

speed so as to respect safety distances from the preceding vehicles. In order to greatly simplify the detection task, the approach employs vehicles bearing visual marks (the rear left and right lamps and a roof lamp). From the known configuration of such visual elements, the targets can be easily located in 3-D [40].

On the other hand, the processing of nonmonocular image sets requires identifying correspondences between pixels in the different images: two images, in the case of stereo vision, and three images, in the case of trinocular vision. The advantage of analyzing stereo images instead of a monocular sequence lies in the possibility of directly detecting the presence of obstacles, which, in the case of an optical flow-based approach, is indirectly derived from the analysis of the velocity field. Moreover, in a limit condition where both vehicle and obstacles have small or null speeds, the optical flow-based approach fails while the other can still detect obstacles.

The Urban Traffic Assistant (UTA) project of the Daimler-Benz research group, for example, aims at an intelligent stop and go for inner-city traffic using stereo vision and obtaining 3-D information in real time. In addition, the UTA demonstrator is able to recognize traffic signs, traffic lights, and walking pedestrians as well as the lane, zebra crossings, and stop lines [21].

Also, the Massachusetts Institute of Technology group developed a cost-effective stereo vision system. The system is used for 3-D lane detection and traffic monitoring, as well as for other on-vehicle applications. The system is able to separate partially overlapped vehicles and distinguish them from shadows.

As another example, the system developed by the Research & Development Center at Toshiba Corporation is based on a domain-specific stereo method for 2-D navigation without depth search and metric camera calibration. Under the assumption that the vehicle is moving on a flat plane, it uses a "pseudo-projective camera model," which provides a good approximation to the general camera model in road scenes [41].

In stereovision, the correct identification of correspondences in the two images represents an important problem. In fact, the size of the areas where the search for corresponding pixels is performed can deeply influence the quality of the results obtained. If the window size is too small, the right match may be missed. On the other hand, if the window size is too large, too many possibilities may exist. To overcome this problem, some systems try to identify the image zones in which it is more probable to find homologous points. The stereo matching algorithm developed at Tohoku University, for example, is based on sum of absolute differences (SAD) computation and uses variable window sizes for each pixel in the image. Window dimensions for a specific pixel are determined by searching for minimums in the corresponding SAD graph [42].

Other systems face the stereo correspondence problem in completely different manners. For instance, the algorithm studied at the Universitè des Sciences et Technologies de Lille is based on a genetic approach. The stereo matching problem is turned into an optimization task where the function representing the constraints of the solution is to be minimized [43].

Furthermore, to decrease the intrinsic complexity of stereo vision, some domain specific constraints are generally adopted.

In the GOLD system, the removal of the perspective effect from stereo images allows to obtain two images that differ only where the initial assumption of a flat road is not valid, thus detecting the free space in front of the vehicle [32].

Analogously, the University of California research unit developed an algorithm that remaps the left image using the point of view of the right image, thus detecting disparities in correspondence to the obstacles. A Kalman filter is then used to track obstacles [44].

Table 2 compares the strong and weak points of the different approaches to the obstacle detection problem.

As mentioned above, a great deal of different techniques have been proposed in the literature and tested on a number of vehicle prototypes in order to solve the Road Following problem, but only a few of them provide an integrated solution (e.g., lane detection and obstacle detection) which, obviously, leads to both an improved quality of the results and to faster and more efficient processing.

The research group of the Istituto Elettrotecnico Nazionale "G. Ferraris" began with limiting the processing to the image portion that is assumed to represent the road, thus relying on the previously discussed lane detection module. This area of the image was analyzed and borders that could represent a potential vehicle were looked for and examined [31].

Moreover, there are situations where the combination of lane detection and obstacle detection is mandatory. This is true, for example, for those systems which focus on the analysis of the vehicle's rear view, with the aim of increasing driver and passenger safety. Without a knowledge of lane structure, in fact, it would be very difficult (if not impossible) to estimate the exact positions of the following vehicles.

Some researchers focus on the implementation of electronic rear-view mirrors, which assist the driver in analyzing what occurs on the road behind him or her. As an example, a system developed at the University of Amsterdam uses a single camera to derive the real-world motion of the vehicles behind the car. Information which can be drawn includes time to contact (to avoid bumper-to-bumper crashes) and lane shifts (useful during overtakings) [45]. The system implemented by Daimler Chrysler AG Research Institute and the Universität der Magdeburg, instead, detects vehicles in the rear view of the host car by means of two cameras, thus exploiting stereovision. Using the steering angle and the detected obstacles, the trajectory of the car can be properly reconstructed and used as a lane change assistant [46].

C. Pedestrian Detection

Vision-based pedestrian detection in outdoor scenes is still an open challenge. People dress in very different colors that sometimes blend with the background, they wear hats or carry bags, and stand, walk and change direction unpredictably. The background is various, containing buildings, moving or parked cars, cycles, street signs, signals etc. Moreover, sudden changes of background are inevitable in vision systems mounted on a moving vehicle.

Many different approaches have been developed to address this complexity. Pattern analysis, stereo vision, shape detection and tracking have been fused in more than one combination. Only few of these systems have already proved their efficacy in applications to intelligent vehicles. Nonetheless, all the principal trends in research will be discussed below to give a broad view on this highly developing field.

The most common approach to pedestrian detection consists of two conceptual steps. First, the image is segmented into foreground and background regions. Then, a second step determines if a foreground region is a pedestrian or not. Most of the work concerning the detection of human shape in cluttered scenes is due to studies on automatic surveillance systems. The fundamental assumption of this research field is a fixed or slow moving camera. In such a situation, a common approach to the detection of regions of interest is to subtract each single frame form a reference frame or an intensity model of the empty field of view built at initialization.

Although very effective, this premise does not fit the requirements of a system for automotive applications, where the background is continuously changing and no modeling is reasonably achievable. Alternative ways of segmentation have been employed, in particular those involving the analysis of more than one image, such as:

- the analysis of *motion*;
- the processing of *stereo images*.

Motion is a common cue to detect interesting regions in a scene. It heavily uses temporal information and has proved to be quite reliable if one wants only to find a moving object and not its precise velocity. Unfortunately, it does not detect standing pedestrians and needs the analysis of a sequence of a few frames before giving a response.

A few works use motion detection with optical flow as a means of segmentation. The basic idea is to detect blobs with a given shape or a common feature, like color, that have similar values of optical flow and track their movement in subsequent frames. A group at the University of Rochester [47] analyzes the scene with a discrete XYT cube, where they assign to each region its average optical flow. An XYT cube is a representation of a sequence of T frames, each divided in $X \times Y$ areas in its two spatial dimensions. In this system, four divisions are used in each spatial dimension, and six along the temporal dimension, resulting in a feature vector of size 96 containing the average optical flow of each region. They then employ Fourier analysis to classify these values. Their method has been applied to the monitoring of some repetitive human activities with a stationary camera, like walking or running on a on gymnastic rolling belt. Like other methods based on optical flow, a good cancellation of ego motion is critical in applications with a moving camera.

Some groups suggest alternative ways for motion detection. A system devised for surveillance applications by the Queen Mary and Westfield College, London [48] uses a zero-crossing detection algorithm using the convolution of a spatio-temporal Gaussian with the history over the values of a pixel in the past six frames. This gave good results, with the extension of a second-order Kalman filter that copes with occlusions. Oiginal work by Cutler and Davis of the University of Maryland [49] uses a subtraction between an image at time t and a version of the same image stabilized with respect to image at instant t-τ. This operation, followed by an appropriate thresholding, gives a map of the pixels representing moving objects. Their method is then based on a two-step approach where recognition is made through analysis of the correlation of two frames taken with a delay of τ. The authors report good performance both with stationary cameras and moving cameras, provided that the background is homogeneous to some extent.

A different approach to the segmentation problem is range thresholding based on *stereo analysis*. In their Bus Driver Assistance project, Zhao and Thorpe of Carnegie Mellon University [50] use range as a means of object segmentation.

Range stereo systems rely on some kind of correlation between left and right images of the same scene and are strongly affected by noise, above all at longer ranges. The authors report problems in the detection procedures due to the fact that each segmented region does not always correspond to a single object. An hypothesis and verification procedure is then necessary to split or group segmented regions, which are then passed to a pedestrian recognition module.

In surveillance applications, stereo analysis is sometimes used as a cue to build a disparity map of the background for use with background subtraction. This is what happens in the system realized by SRI International using their integrated stereo cameras called Small Vision System (SVS) [51]. Segmented objects are therefore organized in pyramids to compensate for scale differences. This system shows low sensibility to distracting elements like shadows, lighting changes, occluding objects or camera dynamics. Moreover, foreground regions may be separated even if they are at the same distance as some background features.

Some systems substitute the segmentation step with a focus-of-attention approach, where salient regions in opportune feature maps are interpreted as candidates for pedestrians. In the GOLD system [52], vertical symmetries are associated with candidates for standing pedestrians, both moving and stationary. Further information derives from symmetry maps of horizontal edges and of their number per column. Then a bounding box encloses interesting regions for a separated recognition step. A more complicated system has been developed at the Ruhr-Universität Bochum [53]. The focus of attention is directed by a composition of a map of the local image entropy, a model-matching module with the shape of a Λ representing human legs, and finally inverse perspective mapping (binocular vision) for the short distance field. This information is combined in a temporal dynamic activation field (DAF) that efficiently allocates computational resources in the following recognition and tracking step.

For what concerns the recognition phase, two main trends are pursued in recent research:

- detection of the typical periodicity of *human gait* in the movement of foreground regions;
- *shape analysis* of foreground regions.

Methods based on gait recognition show a higher robustness, but they require the analysis of multiple frames and easily apply only to pedestrians crossing the street in the path of the vehicle, where the alternating movement of legs is more evident. An important drawback of this family of systems is their inability to correctly classify still persons as pedestrians. On the other hand, shape-based approaches are more sensible to false positives and thus they need a good detection phase, but they correctly recognize even stationary people.

Periodicity of the *human gait* is often recognized with traditional methods like the Fourier transform. Some systems perform a frequency analysis of the changes of candidate patterns over time and then select those that show the frequency spectrum characteristic of human gait. As an example, Cutler and Davis [49] use a short-time Fourier transform with a Hanning windowing function to analyze the signals obtained by correlation of the pattern of detected objects.

In one of the studies for the development of the UTA [54], an Adaptable Time Delay Neural Network (ATDNN) algorithm is considered. After a first stereo-based segmentation that detects and extracts image regions containing legs of pedestrians, the ATDNN performs a local spatio-temporal processing to detect the typical pattern of the movement. This way, the gait patterns of a pedestrian in a complete gait cycle are learned by the network.

In the algorithm developed by the group at Ruhr-Universität Bochum [53], the torso of a candidate pedestrian is tracked so that the lower part of the region can be analyzed to reveal the relative motion of legs. A rough model of two legs consisting of two rod-like pieces each, jointed at the knees, is juxtaposed on the image area below the tracked torso. The periodic movement detected is then correlated to an experimental curve derived from the statistical average of human gait periods. High peaks of the correlation function indicate the presence of a person.

Basic *shape analysis* methods consist in matching a significant and simple shape onto candidate foreground regions. Some systems, like GOLD [52] or the one by SRI International [51], employ a Ω model for the head and shoulders. This approach is very sensible to scale variation, so multiple models of different scales are needed. In the two systems above, three and five different models are used, from coarse to fine resolution, according to the estimated distance of a subject. A group at The Robotics Institute of Carnegie Mellon University [55] uses a skeletonization procedure to characterize the shape of a foreground object previously detected. For each object, they calculate first the centroid of the area and then the distances from the centroid to each border points. Local maxima of the distance function are taken as the external points of the skeleton. The authors suggest that the relative position of centroid and external points, and their rigidity, may be applied to recognition of different types of targets. For what concerns humans, they further confirm the analysis with gait detection.

Another algorithm developed within the UTA project at DaimlerChrysler [56] presents a two-step approach where both phases rely on shape and pattern analysis. The detection step is based on a search of the image with a numerous set of silhouettes of the human body using a distance transform of the edge image. The silhouettes are organized hierarchically with a coarse-to-fine approach, so that generic forms are tried first, and similar and more detailed shapes afterwards. The validation step is accomplished by a radial-basis-function classifier trained with rectangular regions containing pedestrians which were previously selected by a human operator. In a work by the University of Maryland, the system was further improved to perform tracking [57]. A statistical shape model of a pedestrian is first built and then approximated by a linear point distribution model. The tracking of this model over the image sequence is accomplished with a quasi-random sampling method, based on a zero-order motion model with large process noise high enough to account for the greatest expected change in shape

Table 3
Pros and Cons of the Most Typical Assumptions in Lane Detection

	Pros	Cons
Segmentation with motion	Good reliability if background is still or changing slowly, gives strong information for gait detection	Cancellation of ego-motion is needed, needs analysis of a sequence of images
Segmentation with stereo	Low sensibility to lighting changes or occlusions, feasible even with changing backgrounds	Does not segment single objects (hypothesis and verification procedure is needed), unreliable at longer ranges
Focus of attention	Analysis carried out on a single-image basis, fast processing	Features must be carefully selected, does not really segment an object from the background
Recognition of human gait	Gait frequency is highly recognizable, high reliability	Does not recognize standing pedestrians, limited to pedestrians crossing the path of the vehicle, needs analysis of a sequence of images
Recognition of human shape	More versatile, feasible on a single image, fast processing	Problems at multiple scales, human shape has a high variability, difficult if segmentation is not precise

and motion. The authors report a high rate of success and the ability of the tracker to quickly recover from failures.

More systems employ pattern recognition with classifiers to accomplish the recognition step. Sometimes the original imaged is processed before the application of the classifier. For example, Zhao and Thorpe [50] propose a three-layer feedforward network processing the intensity gradient image rather than the original image.

Table 3 summarizes the strong and weak points of the different approaches to the pedestrian detection problem.

The system devised at the AI Lab of MIT [58] for automotive applications fuses the detection and validation steps into one. The image is initially transformed with Haar wavelets and then scanned to detect the pattern associated with a human person. The human pattern is learned, and subsequently recognized, through statistical reasoning with a support vector machine—a technique to train classifiers which is capable of learning in sparse, high-dimensional spaces with very few examples. The system uses multiple classifiers for arms, head, and legs, in a hierarchical organization, in order to cope with occlusions. In [59], the system was adapted to consider temporal information in the form of a joint analysis of five sequential frames.

IV. ARCHITECTURAL ISSUES

In the early years of ITS applications, a great deal of custom solutions were proposed, based on ad hoc, special-purpose hardware. This recurrent choice was motivated by the fact that the hardware available on the market at a reasonably low cost was not powerful enough to provide real-time image processing capabilities. As an example, the researchers of the Universität der Bundeswehr developed their own system architecture: several special-purpose boards were included in the Transputer-based architecture of the VITA vehicle [60].

Others developed or acquired *ad hoc* processing engines based on SIMD computational paradigms to exploit the spatial parallelism of images. Among them, the cases of the 16k Mas-Par MP-2 installed on the experimental vehicle NavLab I [61], [62] at Carnegie Mellon University and the massively parallel architecture PAPRICA [63] jointly developed by the University of Parma and the Politecnico di Torino and tested on the MOB-LAB vehicle.

Besides selecting the proper sensors and developing specific algorithms, a large percentage of this first research stage was therefore dedicated to the design, implementation, and test of new hardware platforms. In fact, when a new computer architecture is built, not only do the hardware and architectural aspects—such as instruction set, I/O interconnections, or computational paradigm—need to be considered, but software issues as well. Low-level basic libraries must be developed and tested along with specific tools for code generation, optimization and debugging.

In the last few years, the technological evolution led to a change: almost all research groups are shifting toward the use of off-the-shelf components for their systems. In fact, commercial hardware has nowadays reached a low price/performance ratio. As an example, both the NavLab 5 vehicle from Carnegie Mellon and the ARGO vehicle from the University of Parma are presently driven by systems based on general-purpose processors. Thanks to the current availability of fast internetworking facilities, even some MIMD solutions are being explored, composed of a rather small number of powerful, independent processors, as in the case of the VaMoRs-P vehicle of the Universität der Bundeswehr on which the Transputer processing system has now been partly replaced by a cluster of three PCs (dual Pentium II) connected via a fast ethernet-based network [20].

Current trends, however, are moving toward a mixed architecture, in which a powerful general-purpose processor is aided by specific hardware such as boards and chips implementing optical flow computation, pattern-matching, convolution, and morphological filters. Moreover, some SIMD capabilities are now being transferred into the instruction set of the last-generation CPUs, which has been tailored to exploit the parallelism intrinsic to the processing of visual and audio (multimedia) data. The MMX extensions of the Intel Pentium processor, for instance, are exploited by the GOLD system which acts as the automatic driver of the ARGO vehicle to boost up performance.

In conclusion, it is important to emphasize that, although the new generation of systems are all based on commercial hardware, the development of custom hardware has not lost significance, but is gaining a renewed interest for the production of embedded systems. Once a hardware and software prototype has been built and extensively tested, its functionalities have to be integrated in a fully optimized and engineered embedded system before marketing. It is in this stage of the project that the development of ad hoc custom hardware still plays a fundamental role and its costs are justified through a large scale market.

V. Perspectives on Intelligent Vehicles

The promising results obtained in the first stages of research on intelligent vehicles demonstrate that a full automation of traffic (at least on motorways or sufficiently structured roads) is technically feasible.

Nevertheless, besides technical problems, some issues must be carefully considered in the design of these systems such as legal aspects related to the responsibility in case of faults and incorrect behavior of the system, and the impact of automatic driving on human passengers. User acceptance in particular will play a critical role in how intelligent vehicles will look and perform and the system interface will have a strong influence on how a user will view and understand the functionality of the system.

Therefore, a long period of exhaustive tests and refinement must precede the availability of these systems on the general market, and a fully automated highway system with intelligent vehicles driving and exchanging information is not expected for a couple of decades.

For the time being, complete automation will be restricted to special infrastructures such as industrial applications or public transportation. Then, automatic vehicular technology will be gradually extended to other key transportation areas such as goods shipping, for example on expensive trucks, where the cost of an autopilot is negligible with respect to the cost of the vehicle itself and the service it provides. Finally, once technology has stabilized and the most promising solution and best algorithms fixed, a massive integration and a widespread use of such systems will take place in private vehicles, but this will not happen for another two or more decades.

References

[1] C. Little, "The intelligent vehicle initiative: Advancing 'Human-Centered' smart vehicles," *Public Roads Mag.*, vol. 61, no. 2, pp. 18–25, Sept./Oct. 1997.
[2] M. Cellario, "Human-Centered intelligent vehicles: Toward multimodal interface integration," *IEEE Intell. Syst.*, vol. 16, no. 4, pp. 78–81, July/Aug. 2001.
[3] "Vehicle-highway automation activities in the United States," U.S. Dept. of Transportation, 1997.
[4] H. Tokuyama, "Asia-pacific projects status and plans," U.S. Dept. of Transportation.
[5] M. C. Hulse *et al.*, "Development of human factors guidelines for advanced traveler information systems and commercial vehicle operations: Identification of the strengths and weaknesses of alternative information display formats," Federal Highway Administration, Washington, DC, Tech. Rep. FHWA-RD-96-142, 1998.
[6] U. Seger, H. G. Graf, and M. E. Landgraf, "Vision assistance in scenes with extreme contrast," *IEEE Micro*, vol. 13, pp. 50–56, Jan.–Feb. 1993.
[7] C. G. Sodini and S. J. Decker, "A 256 × 256 CMOS brightness adaptive imaging array with column-parallel digital output," in *Proc. IEEE Int. Conf. Intelligent Vehicles*, 1998, pp. 347–352.
[8] M. Mizuno, K. Yamada, T. Nakano, and S. Yamamoto, "Robustness of lane mark detection with wide dynamic range vision sensor," in *Proc. IEEE Int. Conf. Intelligent Vehicles*, 1995, pp. 171–176.
[9] T. M. Jochem, D. A. Pomerleau, and C. E. Thorpe, "MANIAC: A next generation neurally based autonomous road follower," in *Proc. 3rd Int. Conf. Intelligent Autonomous Systems*, 1993.
[10] J. D. Crisman and C. E. Thorpe, "UNSCARF, A color vision system for the detection of unstructured roads," *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 2496–2501, 1991.
[11] K. I. Kim, S. Y. Oh, S. W. Kim, H. Jeong, C. N. Lee, B. S. Kim, and C. S. Kim, "An autonomous land vehicle PRV II: Progresses and performance enhancement," *Proc. IEEE IV*, pp. 264–269, 1995.
[12] A. Takahashi, Y. Ninomiya, M. Ohta, and K. Tange, "A robust lane detection using real-time voting processor," in *Proc. IEEE ITS*, 1999, pp. 577–580.
[13] S. L. Michael Beuvais and C. Kreucher, "Building world model for mobile platforms using heterogeneous sensors fusion and temporal analysis," *Proc. IEEE ITS*, p. 101, 1997.
[14] A. Coda, P. C. Antonello, and B. Peters, "Technical and human factor aspects of automatic vehicle control in emergency situations," in *Proc. IEEE ITS*, 1997.
[15] D. A. Pomerleau and T. Jochem, "Rapidly adapting machine vision for automated vehicle steering," *IEEE Expert*, vol. 11, Apr. 1996.
[16] J. P. Gonzàlez and Ü. Özgüner, "Lane detection using histogram-based segmentation and decision trees," in *Proc. IEEE ITS*, 2000, pp. 346–351.
[17] P. Charbonnier, P. Nicolle, Y. Guillard, and J. Charrier, "Road boundaries detection using color saturation," in *Proc. 9th Eur. Signal Processing Conf.'98*, Sept. 1998.
[18] R. Chapuis, R. Aufrère, F. Chausse, and J. Alizon, "Road sides recognition under unfriendly lighting conditions," in *Proc. IEEE IV*, 2001, pp. 13–18.
[19] J. Goldbeck, D. Graeder, B. Huertgen, S. Ernst, and F. Wilms, "Lane following combining vision and DGPS," in *Proc. IEEE IV*, 1998, pp. 445–450.
[20] M. Lützeler and E. D. Dickmanns, "Road recognition with MarVEye," in *Proc. IEEE IV*, 1998, pp. 341–346.
[21] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler, "Autonomous driving goes downtown," in *Proc. IEEE IV*, 1998, pp. 40–48.
[22] J. Goldbeck and B. Huertgen, "Lane detection and tracking by video sensors," in *Proc. ITS*, 1999, pp. 74–79.
[23] K. A. Redmill, S. Upadhya, A. Krishnamurthy, and Ü. Özgüner, "A lane tracking system for intelligent vehicle applications," in *Proc. IEEE ITS*, 2001, pp. 275–281.
[24] K. A. Redmill, "A simple vision system for lane keeping," *Proc. IEEE ITS*, 1997.

[25] F. Chausse, R. Aufrère, and R. Chapuis, "Vision based vehicle trajectory supervision," in *Proc. IEEE ITS*, 2000, pp. 143–148.

[26] J. Goldbeck and B. Huertgen, "Lane detection and tracking by video sensors," in *Proc. ITS*, 1999, pp. 74–79.

[27] R. Risack, P. Klausmann, W. Kruger, and W. Enkelmann, "Robust lane recognition embedded in a real-time driver assistance system," in *Proc. IEEE IV*, 1998, pp. 35–40.

[28] S. M. Wong and M. Xie, "Lane geometry detection for the guidance of smart vehicle," in *Proc. IEEE ITS*, 1999, pp. 925–928.

[29] X. Youchun, W. Rongben, and J. Shouwen, "A vision navigation algorithm based on linear lane model," in *Proc. IEEE IV*, 2000, pp. 240–245.

[30] A. Broggi and S. Bertè, "Vision-based road detection in automotive systems: A real-time expectation-driven approach," *J. Artif. Intell. Res.*, vol. 3, pp. 325–348, Dec. 1995.

[31] S. Denasi, C. Lanzone, P. Martinese, G. Pettiti, G. Quaglia, and L. Viglione, "Real-time system for road following and obstacle detection," in *Proc. SPIE on Machine Vision Applications, Architectures, and Systems Integration III*, vol. 2347, 1994, pp. 70–79.

[32] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Processing*, vol. 7, pp. 62–81, Jan. 1998.

[33] R. Aufrère, R. Chapuis, and F. Chausse, "A fast and robust vision based road following algorithm," in *Proc. IEEE IV*, 2000, pp. 192–197.

[34] A. Broggi, M. Bertozzi, A. Fascioli, and G. Conte, "Automatic vehicle guidance: The experience of the ARGO vehicle," *World Scientific*, 1999.

[35] S. Kyo, T. Koga, K. Sakurai, and S. Okazaki, "A robust vehicle detecting and tracking system for wet weather conditions using the IMAP-VISION image processing board," in *Proc. IEEE ITS*, 1999, pp. 423–428.

[36] S. Denasi and G. Quaglia, "Obstacle detection using a deformable model of vehicles," in *Proc. IEEE IV*, 2001, pp. 145–150.

[37] W. Kruger, W. Enkelmann, and S. Rossle, "Real-time estimation and tracking of optical flow vectors for obstacle detection," *Proc. IEEE IV*, pp. 304–309, 1995.

[38] S. M. Smith and J. M. Brady, "ASSET-2: Real-time motion segmentation and shape tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 814–892, Aug. 1995.

[39] Z. Hu and K. Uchimura, "Tracking cycle: A new concept for simultaneously tracking of multiple moving objects in a typical traffic scene," in *Proc. IEEE IV*, 2000, pp. 233–239.

[40] F. Marmoiton, F. Collange, and J. P. Dèrutin, "Location and relative speed estimation of vehicles by monocular vision," in *Proc. IEEE IV*, 2000, pp. 227–232.

[41] H. Hattori, "Stereo for 2D visual navigation," in *Proc. IEEE IV*, 2000, pp. 31–38.

[42] M. Hariyama, T. Takeuchi, and M. Kameyama, "Reliable stereo matching for higly-safe intelligent vehicles and its VLSI implementation," in *Proc. IEEE IV*, 2000, pp. 128–132.

[43] Y. Ruichek, H. Issa, and J. Postaire, "Genetic approach for obstacle detection using linear stereo vision," in *Proc. IEEE IV*, 2000, pp. 261–266.

[44] D. Koller, J. Malik, Q.-T. Luong, and J. Weber, "An integrated stereo-based approach to automatic vehicle guidance," in *Proc. 5th Int. Conf. Computer Vision*, 1995, pp. 12–20.

[45] M. B. van Leeuwen and F. C. A. Groen, ""Motion estimation with a mobile camera for traffic applications," in *Proc. IEEE IV*, 2000, pp. 58–63.

[46] C. Knoeppel, A. Schanz, and B. Michaelis, "Robust vehicle detection at large distance using low resolution cameras," *Proc. IEEE IV*, pp. 267–272, 2000.

[47] R. Polana and R. C. Nelson, "Detection and recognition of periodic, nonrigid motion," *Int. J. Comp. Vis.*, vol. 23, no. 3, pp. 261–282, 1997.

[48] S. J. McKenna and S. Gong, "Non-intrusive person authentication for access control by visual tracking and face recognition," in *Int. Conf. Audio and Video Authentication*, 1997, pp. 177–184.

[49] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 781–796, Aug. 2000.

[50] L. Zhao and C. Thorpe, "Stereo and neural network based pedestrian detection," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 3, pp. 148–154, 1999.

[51] D. Beymer and K. Konolige, "Real-time tracking of multiple people using continuous detection," in *Proc. Int. Conf. Comp. Vis.*, 1999.

[52] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proc. IEEE IV*, 2000, pp. 215–220.

[53] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking pedestrian recognition," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, no. 3, pp. 155–163, 2000.

[54] C. Wöhler, U. Kressler, and J. K. Anlauf, "Pedestrian recognition by classification of image sequences. Global approaches vs local spatio-temporal processing," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2000.

[55] H. Fujiyoshi and A. Lipton, "Real-time human motion analysis by image skeletonization," in *Proc. IEEE WACV'98*, 1998, pp. 15–21.

[56] D. M. Gavrila, "Pedestrian detection from a moving vehicle," in *Proc. Eur. Conf. Comp. Vis.*, 2000, pp. 37–49.

[57] V. Philomin, R. Duraiswami, and L. Davis, "Pedestrian tracking from a moving vehicle," in *Proc. IEEE IV*, 2000, pp. 350–355.

[58] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 4, pp. 349–361, 2001.

[59] C. Papageorgiou and T. Poggio, "A pattern classification approach to dynamical object detection," in *Int. Conf. Computer Vision*, 1999, pp. 1223–1228.

[60] E. D. Dickmanns, "Expectation-based multi-focal vision for vehicle guidance," in *Proc. 8th Eur. Signal Processing Conf.*, 1995, pp. 1023–1026.

[61] T. M. Jochem and S. Baluja, "A massively parallel road follower," in *Proc. IEEE Computer Architectures for Machine Perception*, M. A. Bayoumi, L. S. Davis, and K. P. Valavanis, Eds., 1998, pp. 2–12.

[62] ——, "Massively parallel, adaptive, color image processing for autonomous road following," in *Massively Parallel Artificial Intelligence*, H. Kitano, Ed. Cambridge, MA: MIT Press, 1993.

[63] A. Broggi, G. Conte, F. Gregoretti, C. Sansoè, and L. M. Reyneri, "The evolution of the PAPRICA system," *Integr. Computer-Aided Eng. J.*, vol. 4, no. 2, pp. 114–136, 1997.

Massimo Bertozzi (Associate Member, IEEE) received the Dr.Eng. (Master) degree in electronic engineering and the Ph.D. degree in information technology, both from the Università di Parma, Italy, in 1994 and 1997, respectively. His master's thesis was on the implementation of simulation of Petri nets on the CM-2 massive parallel architecture, while his Ph.D. dissertation was on real-time image processing for automotive applications.

From 1994 to 1997, he chaired the Parma University IEEE student branch. His research interests focus mainly on the application of image processing to real-time systems and to vehicle guidance, the optimization of machine code at assembly level, and parallel and distributed computing. He is currently an Associate Researcher in the Dipartimento di Ingegneria dell'Informazione, Università di Parma.

Alberto Broggi (Associate Member, IEEE) received the Dr.Eng. (Master) degree in electronic engineering and the Ph.D. degree in information technology both from the Università di Parma, Italy, in 1990 and 1994, respectively.

From 1994 to 1998, he was an Associate Researcher at the Dipartimento di Ingegneria dell'Informazione, Università di Parma, Italy. From 1998 to 2001, he was an Associate Professor of Artificial Intelligence at the Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy, and, since 2001, he has been Professor of Computer Science at the University of Parma, Italy. His research interests include real-time computer vision approaches for the navigation of unmanned vehicles, and the development of low-cost computer systems to be used on autonomous agents. He is the Coordinator of the ARGO project, with the aim of designing, developing, and testing the ARGO autonomous prototype vehicle, equipped with special active safety features and enhanced driving capabilities. He is the author of more than 120 refereed publications in international journals, book chapters, and conference proceedings. He is actively involved in the organization of scientific events and is on the Editorial Board and Program Committee of many international journals and conferences and has been invited to act as Guest Editor of journals and magazines theme issues on topics related to intelligent vehicles, computer vision application, and computer architectures for real-time image processing.

Massimo Cellario received the Dr.Eng. (Master) degree in information engineering from the University of Pavia, Pavia, Italy, in 1998, and the Ph.D. degree in computer science in 2002. His thesis concerned nonimmersive virtual environments and his dissertation focused on human-centered interface intergration in intelligent vehicles.

His research interests focus on multi-modal/perceptual interfaces for human–computer interaction, three-dimensional computer vision, and image synthesis.

Paolo Lombardi received the B.E. and M.Sc. degrees in electronic engineering form the University of Pavia, Italy, in 2000. He is currently working toward the Ph.D. degree in electronic engineering and computer science within a program of cotutoring between the University of Pavia, Italy, and the University of Paris Sud, France.

His research interests include computer vision with applications to automotive and visual attention algorithms.

Alessandra Fascioli (Member, IEEE) received the Dr.Eng. (Master) degree in electronic engineering from the Università di Parma, Italy, in 1996, and her PhD. degree in 2000. Her thesis focused on stereo vision-based obstacle localization in automotive environments.

From November 1996 to October 1999, she was a Ph.D. student in information technology at the Dipartimento di Ingegneria dell'Informazione, University of Parma, where she chaired the local IEEE student branch. She is currently a temporary researcher at the University of Parma. Her research interests focus on real-time computer vision and computer architectures for automatic vehicle guidance. She is also interested in image processing techniques based on the Mathematical Morphology computational model.

Ms. Fascioli is a member of the IEEE Computer Society, AI*IA, and IAPR.

Marco Porta received the Dr.Eng. (Master) degree in electronic engineering from the Politecnico di Milano, Milan, Italy, in 1996 and the Ph.D. degree in electronic and computer engineering from the Università di Pavia, Pavia, Italy, in 1999. His thesis discussed a system for remote control of an autonomous mobile robot. His dissertation focused on visual programming languages and their use to achieve complex functionalities with little effort.

Since January 2000, he has been a Post-Doctoral Researcher at the Università di Pavia. His interests include visual languages, multimedia interfaces and vision-based techniques for human–computer interaction (perceptive user interfaces).