Survey of Pedestrian Detection for Advanced Driver Assistance Systems

Article *in* IEEE Transactions on Software Engineering · July 2010

DOI: 10.1109/TPAMI.2009.122 · Source: PubMed

4 authors, including:

David Geronimo
Catchoom
37 PUBLICATIONS **1,310** CITATIONS
SEE PROFILE

Antonio M. López
Autonomous University of Barcelona
207 PUBLICATIONS **4,999** CITATIONS
SEE PROFILE

Angel Domingo Sappa
Autonomous University of Barcelona
173 PUBLICATIONS **2,146** CITATIONS
SEE PROFILE

Survey of Pedestrian Detection for Advanced Driver Assistance Systems

David Gerónimo, Antonio M. López, Angel D. Sappa, *Member*, *IEEE*, and Thorsten Graf

Abstract—Advanced driver assistance systems (ADASs), and particularly pedestrian protection systems (PPSs), have become an active research area aimed at improving traffic safety. The major challenge of PPSs is the development of reliable on-board pedestrian detection systems. Due to the varying appearance of pedestrians (e.g., different clothes, changing size, aspect ratio, and dynamic shape) and the unstructured environment, it is very difficult to cope with the demanded robustness of this kind of system. Two problems arising in this research area are the lack of public benchmarks and the difficulty in reproducing many of the proposed methods, which makes it difficult to compare the approaches. As a result, surveying the literature by enumerating the proposals one-after-another is not the most useful way to provide a comparative point of view. Accordingly, we present a more convenient strategy to survey the different approaches. We divide the problem of detecting pedestrians from images into different processing steps, each with attached responsibilities. Then, the different proposed methods are analyzed and classified with respect to each processing stage, favoring a comparative viewpoint. Finally, discussion of the important topics is presented, putting special emphasis on the future needs and challenges.

Index Terms—ADAS, pedestrian detection, on-board vision, survey.

1 Introduction

Due to the rise in the popularity of automobiles over the last century, road accidents have become an important cause of fatalities. About 10 million people become traffic casualties around the world each year, and two to three million of these people are seriously injured [1], [2]. For instance, in 2003, the United Nations reported almost 150,000 injured and 7,000 killed in vehicle-to-pedestrian accidents just in the European Union alone [3].

Both the scientific community and the automobile industry have contributed to the development of different types of protection systems in order to improve traffic safety. Initially, improvements consisted of simple mechanisms like seat belts, but then more complex devices, such as *antilock bracking systems*, *electronic stabilization programs*, and *airbags*, were developed. Over the last decade, research has moved toward more intelligent on-board systems that aim to anticipate accidents in order to avoid them or to mitigate their severity. These systems are referred to as *advanced driver assistance systems* (ADASs) [2], [4], [5], as they assist the driver in marking decisions, provide signals in possibly dangerous driving situations, and execute counteractive measures. Some examples are the *adaptive cruise control*, which maintains a safe gap between vehicles and the *lane departure warning* that acts when the car is driven out of a lane inadvertently.

In this paper, we focus on a particular type of ADAS, *pedestrian protection systems* (PPSs). The objective of a PPS is to detect the presence of both stationary and moving people in a specific area of interest around the moving host vehicle in order to warn the driver, perform braking actions, and deploy external airbags if a collision is unavoidable (evasive actions could be an option if the pedestrian surroundings are sensed). Accident statistics indicate that 70 percent of the people involved in car-to-pedestrian accidents were in front of the vehicle, of which 90 percent were moving [6]. Therefore, PPSs typically use forward-facing sensors.

The main challenges of a PPS involve detection of pedestrians. These challenges are summarized by the following points:

- The appearance of pedestrians exhibits very high variability since they can change pose, wear different clothes, carry different objects, and have a considerable range of sizes (especially in terms of height).
- Pedestrians must be identified in outdoor urban scenarios, i.e., they must be detected in the context of a cluttered background (urban areas are more complex than highways) under a wide range of illumination and weather conditions that vary the quality of the sensed information (e.g., shadows and poor contrast in the visible spectrum). In addition, pedestrians can be partially occluded by common urban elements, such as parked vehicles or street furniture.
- Pedestrians must be identified in highly dynamic scenes since both the pedestrian and camera are in motion, which complicates tracking and movement analysis. Furthermore, pedestrians appear at different viewing angles (e.g., lateral and front/rear

- *D. Gerónimo, A.M. López, and A.D. Sappa are with the Computer Vision Center and the Computer Science Department, Universitat Autònoma de Barcelona, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain. E-mail: {dgeronimo, antonio, asappa}@cvc.uab.es.*
- *T. Graf is with Volkswagen AG, Electronics Research, 38436 Wolsburg, Germany. E-mail: thorsten.graf@volkswagen.de.*

Manuscript received 29 July 2008; revised 31 Dec. 2008; accepted 13 May 2009; published online 21 May 2009.
Recommended for acceptance by B. Schiele.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-07-0448.
Digital Object Identifier no. 10.1109/TPAMI.2009.122.

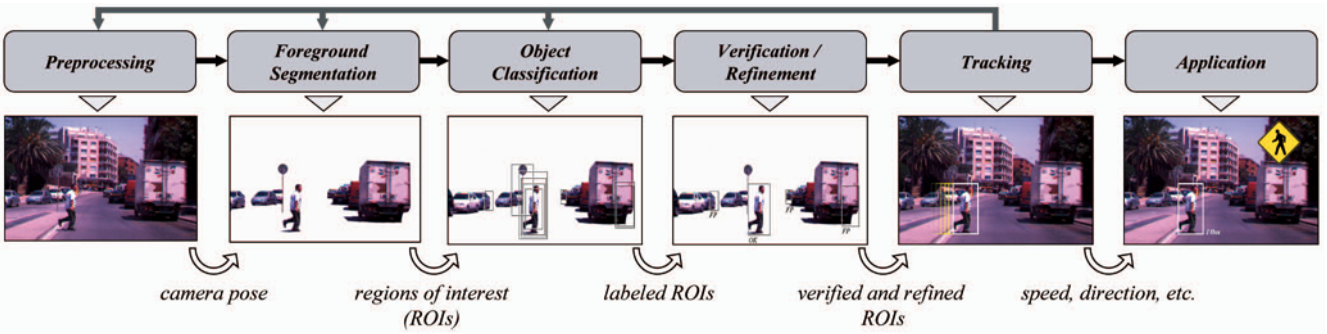

Fig. 1. The architecture proposed for an on-board pedestrian detection system, exemplified for the case of using a camera sensor working in the visible spectrum. The diagram is a simplification that covers the structure of most of the systems, so particular module organizations presented in some papers, for example, interchanging tracking and verification stages, have not been included. However, potential feedback between modules (e.g., tracking-foreground segmentation) is becoming common, so it has been illustrated by the top arrows.

positions) and the system must work over a large range of distances (at least 25 m, which roughly corresponds to a 30×60 pixel pedestrian with a typical 6 mm focal length 640×480 pixel camera).

- The required performance is quite demanding in terms of system reaction time and robustness (i.e., false alarms versus misdetections).

It is clear that the topic differs from general human detection systems, such as surveillance applications or human-machine interfaces, for which some simplifications can be implemented. For example, use of a static camera allows the use of background subtraction techniques.[1]

The first investigations on pedestrian detection for PPSs were presented in the late 1990s. Since then, PPSs have become a hot technological challenge that is of major interest to governments, automotive companies, suppliers, universities, and research centers. As a result, many papers addressing on-board pedestrian detection have been published, a few of which partially survey the state of the art. For instance, in 2001, Gavrila [9] overviewed the few existing systems at that time, focusing on the employed sensors. In 2006, Gandhi and Trivedi [10] followed the same approach, but focused on the aspects of collision prediction and pedestrian behavior analysis. The same authors recently presented a survey that reviews infrastructure developments, sensors, and pedestrian detection approaches in a general transport safety context, rather than focusing especially on on-board detection [11]. Nevertheless, unlike other fields, such as face or vehicle detection, in which in-depth reviews analyzing the algorithms and successful systems have been presented (e.g., [12], [13]), pedestrian detection for ADAS lacks an exhaustive review.

The contribution of this survey is threefold. First, it presents a general module-based architecture that simplifies the comparison of specific detection tasks. The same system breakdown has been successfully used in a short paper [14] to analyze different systems that work in the visible spectrum. Second, it provides a comprehensive up-to-date review of state-of-the-art sensors and benchmarking. Unlike [11], this paper focuses on the techniques used in PPSs, rather than on general pedestrian safety. Moreover, we review different approaches according to the specific tasks defined in the aforementioned architecture, and thus, the description and comparison of each are more detailed. Third, we provide analysis and discussion. Due to the lack of common benchmarks for validation and the complexity of reproducing different approaches, quantitative comparisons become difficult. We present an analysis of the most important proposals in each module and provide quantitative evaluation when possible. In addition, general discussions of the overall systems are also given, pointing out the current limitations and future trends from a more general viewpoint.

The remainder of the paper is organized as follows: In Section 2, we propose a decomposition of PPSs into different processing steps. This architecture is then used as a common framework in which we review the different proposals in the literature, making it easier to understand the requirements, responsibilities, and advantages of the techniques in each module. While most of the approaches use a single type of sensor (camera), some authors propose the fusion of complementary sensors. Such alternative sensors are reviewed at the end of the section. The different techniques used by the most relevant systems are concisely detailed in Tables 2, 3, 4, and 6. Benchmarking, which is a crucial topic in any intelligent system, is explained in Section 3. Discussion of the most important topics for future research, with special emphasis on challenges and needs, is presented in Section 4. Finally, in Section 5, we summarize the aims, content, and conclusions of this paper.

2 Literature Review

The following modules are proposed for splitting the architectures of pedestrian detectors for PPSs, listed according to the processing pipeline order: preprocessing, foreground segmentation, object classification, verification/refinement, tracking, and application. Fig. 1 shows a schematic overview of the modules.

Although some of the proposed modules are not present in the surveyed works and others can be grouped into just one algorithm, we think that most of the systems can be conceptually broken down to fit this architecture for the purpose of comparison. Such a breakdown of any complex system is necessary to provide an ordered analysis of disparate methods. For instance, in Sun's vehicle detection review [13], techniques are divided into

1. We refer the reader to the surveys in [7] and [8] for more details about human detection in applications different other than PPSs.

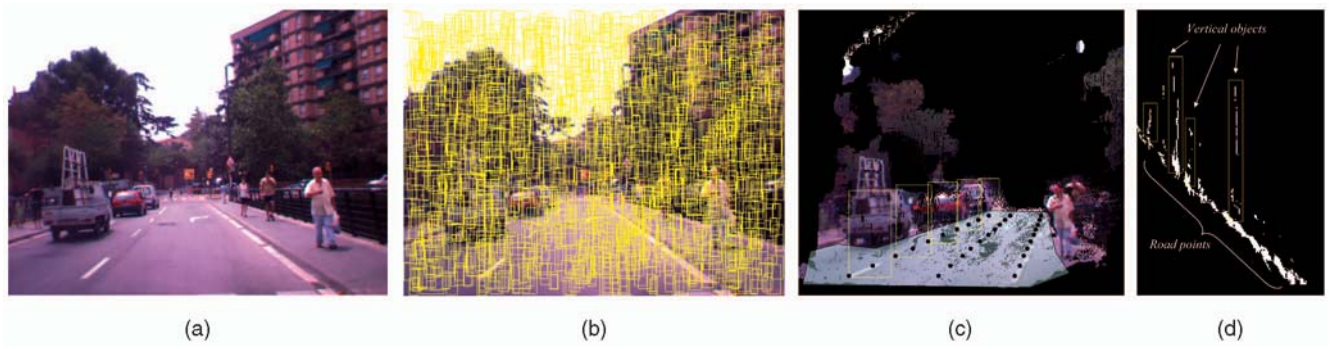

Fig. 2. Foreground segmentation schemes. (a) Original image. (b) Exhaustive scan [31] (just showing 10 percent of the ROIs). (c) Sketch of road scanning after road fitting in euclidean space [30]. (d) Results of v-disparity applied to the same frame [24].

hypothesis generation and hypothesis validation, thus allowing the reader to concentrate on the methods for solving simpler problems, rather than approaching the problem as a whole.

The following sections (Sections 2.1-2.6) describe the mentioned modules, review the existing techniques, and provide some analysis and comparison when possible. In order to provide a sensible comparison of the analyzed approaches, all of the works in these sections make use of passive sensors, i.e., cameras that work in either the visible (typically, for daytime) or infrared (for nighttime) spectra. In fact, they are the most commonly used sensors for PPSs. Henceforth, we will refer to the visible spectrum as VS (i.e., the range 0.4-0.75 μm) and the infrared either as NIR (near infrared, 0.75-1.4 μm) or TIR[2] (thermal infrared, 6-15 μm). The sensibility of NIR sensors ranges from 0.4 to 1.4 μm, so it can be said that they work in the VS+NIR spectrum. Regarding TIR sensors, they capture relative temperature, which is very convenient for distinguishing *hot* targets like pedestrians or vehicles from *cold* ones like asphalt or trees. For the sake of completeness, we include Section 2.7 that describes other sensors and review some systems that exploit the so-called sensor fusion. Finally, Tables 2, 3, 4, 5, and 6 provide some details in a visual comparative manner that have been omitted from the text.

2.1 Preprocessing

The preprocessing module includes tasks such as exposure time, gain adjustments, and camera calibration, to mention a few.

2.1.1 Review

Although low-level adjustments, such as exposure or dynamic range, are normally not described in ADAS literature, some recently published papers have targeted image enhancement through these systems. Real-time adjustments are a recurring difficulty in this area, especially in urban scenarios. For example, short tunnels, narrow streets, and the rapid motion of the scene (common conditions in PPSs) can result in images with over/undersaturated areas or poorly adjusted dynamic range, which creates additional difficulties for the latter algorithms in the system. Although not specifically devoted to ADAS, Nayar and Branzoi [15] present some approaches for performing a locally adaptive dynamic range: fusion of different exposures, spatial filter mosaicing and pixel exposures, multiple image/pixel sensors, etc. Besides, during recent years, solutions exploiting High Dynamic Range (HDR) images [16], [17] have gained interest in driver assistance due to their potential to provide high contrast in the aforementioned scenarios. In fact, HDR cameras cover the VS+NIR spectrum, so they are also useful for nighttime vision.

Camera calibration is another main topic in the processing module. Few approaches tackle both intrinsic and extrinsic on-board self-calibration [18], [19]. The most common approach is to initially compute the intrinsic parameters, and then, to assume that they are constant while the extrinsic parameters are continuously updated. This procedure, which is often referred to as camera pose estimation, avoids the so-called constant road slope assumption, a simplification that is not applicable to real PPSs given the road slope variability in urban scenarios and the changes in vehicle dynamics.

The existing approaches can be divided into two categories: monocular-based and stereo-based. In the former case, the algorithms are mainly based on the study of visual features. In [20], [21], Broggi et al. correct the vertical image position by relying on the detection of horizontal edges oscillations: The horizon line is computed according to the previous frames. A comparative study of different monocular camera pose estimation approaches has been presented in [22]. It included horizontal edges, features-based, and frame difference algorithms. Recently, Hoiem et al. [23] presented a probabilistic framework for 3D geometry estimation based on a monocular system. A training process, based on a set of 60 manually labeled images, is applied to form a prior estimated of the horizon position and camera height (i.e., camera pose values).

Regarding stereo-based pose estimation, Labayrade et al. introduced v-disparity space [24], which consists of accumulating stereo disparity along the image y-axis in order to 1) compute the slope of the road (which is related to the horizon line) and 2) point out the existence of vertical objects when the accumulated disparity of an image row is very different from its neighbors (Fig. 2d). Extensions of this representation can be found in [25]. Other approaches work in euclidean space. For instance, Sappa et al. [26] proposed fitting 3D road data points to a plane, whereas Nedevschi et al. [27] use a clothoid. In the euclidean space, classical least

2. The sensors that we refer to as TIR are sometimes called night vision, thermal infrared, infrared alone, or far infrared (FIR).

squares fitting approaches can be followed, while in the v-disparity space, voting schemes are generally preferred (e.g., Hough transform). Recently, Ess et al. [28], [29] proposed the use of pedestrian location hypotheses together with depth cues to estimate the ground plane, which is used to reinforce new detections tracks. The authors call this approach cognitive feedback, in the sense that a loop is established between the classification and tracking modules and ground plane estimation.

2.1.2 Analysis

HDR sensors provide the possibility of obtaining highly contrasted images in outdoor scenarios. This technology will be of crucial importance in PPSs in order to avoid the over/undersaturated regions that are typically seen in ADAS imaging. In fact, many of the failures of current detection algorithms correspond to poorly contrasted images (see databases in Section 3), so this technology will undoubtedly benefit the system performance.

Stereo-based approaches provide more robust results in camera pose estimation than monocular approaches. Horizon-like stabilizers are based on the assumption that the changes in the scene are smooth, which is not always a valid assumption in urban scenarios. Moreover, in such monocular-based approaches, the global error increases with time as long as the estimation depends on previous frames (the drift problem). On the contrary, stereo-based approaches (both disparity and 3D data) do not accumulate errors and can provide information about the object's distance from the vehicle. It is not clear whether disparity-based approaches are better than 3D data-based approaches. Each approach presents advantages, disadvantages, and limitations. For example, disparity-based approaches are generally faster than those based on 3D data points are; however, they are limited to planar road approximations, while 3D-based approaches allow plane, clothoid, and any free form surface approximation. The more recent of the reviewed works shows a clear trend toward using stereo-based approaches to obtain accurate camera pose estimates in spite of the additional CPU time required for disparity/depth estimation.

2.2 Foreground Segmentation

Foreground Segmentation, which is also referred to as *candidate generation*, extracts *regions of interest* (ROI) from the image to be sent to the classification module, avoiding as many background regions as possible. Although some papers do not contain a specific segmentation module (e.g., exhaustive scanning), these techniques are of remarkable importance not only to reduce the number of candidates, but also to avoid scanning regions like the sky. The key to this stage is to avoid missing pedestrians; otherwise, the subsequent modules will not be able to correct the error. In describing this module, we will often use the term *pedestrian size constraints* (PSCs), which refer to the aspect ratio, size, and position that candidate ROIs must fulfill to be considered to contain a pedestrian (e.g., in [30], pedestrians are assumed to be around 1.70 m, with some standard deviation, e.g., 20 cm, tall with a $1/2$ aspect ratio; hence, ROIs are constrained to these parameters).

2.2.1 Review

The simplest candidate generation procedure is an exhaustive scanning approach [31], [32] that selects all of the possible candidates in an image according to PSC, without explicit segmentation. For instance, in [31], the authors start by scanning the image with ROIs of 64×128 pixels, moving this window in increments of 8 pixels. Then, they reduce the image size by a factor of 1.2 and perform the same scan again. This procedure has two main drawbacks: 1) The number of candidates is large (see Fig. 2b), which makes it difficult to fulfill real-time requirements, although some proposals have recently studied this problem [33], [34], [35], and 2) many irrelevant regions are passed to the next module (e.g., sky regions or ROIs inconsistent with perspective), which increases the potential number of false positives. As a result, other approaches are used to perform explicit segmentation.

2D-based. Miau et al. [36], [37] use a biologically inspired attentional algorithm that selects ROIs according to color, intensity, and gradient orientation of pixels. In several works from *Parma University*, the vertical symmetries in the visible [38], [39], [40], [41] and TIR spectra are used alone [42], [20] or as a complement to stereoimaging [39]. In this case, ROIs are adjusted around each symmetry axis maintaining the PSC. The presence of many horizontal edges is often taken into account as a nonpedestrian quality.

Intensity thresholding is the most intuitive segmentation technique when dealing with TIR images. Some implementations include single thresholding [43], double image and hot-spots-based thresholding [44], and adaptive intensity-based thresholding [45], [46]. Another simple technique is the use of vertical and horizontal histogram projection together with thresholding [47], [48]. *Hypermutation networks* [49], which use a multistage neighborhood pixel classification, are considered in more sophisticated approaches like [50], [51] to classify pixels as foreground/background. In [50], output pixels from the network are grouped using connected component analysis, so the algorithm can be understood as a segmentation/classification process.

Stereo. Franke and Kutzbach [52] present one of the first stereo algorithms specifically developed for ADAS. Local structure classification, which resolves ambiguities by the use of a disparity histogram, is used to perform stereo correspondence. They extended the algorithm with a multiresolution method with subpixel accuracy in [53], [54]. These works become more relevant when they are used in the well-known PROTECTOR system [55]. In this system, the returned map is multiplexed into different discrete depth ranges, which are then scanned with PSC-windows, taking into account the location of the ground plane. If the depth features in one of the windows exceed a given percentage, the window is added to the ROI list supplied to the classifier.

Many authors [38], [56], [57] have made use of the aforementioned v-disparity representation [24] to identify ground and vertical objects. Extraction of candidate regions bounding vertical objects is straightforward after the removal of road surface points. These approaches are based on the fact that a plane (i.e., road surface) in euclidean space

becomes a straight line in the v-disparity space (Fig. 2d). In [30], Gerónimo et al. use stereo-based road plane fitting [26] to dynamically select a set of ROIs that are lying on the ground and satisfying the PSC avoiding the flat road assumption (Fig. 2c). Disparity map analysis together with PSC is also used to extract candidates [58], [39], [59].

Recently, Krotosky and Trivedi proposed the use of multimodal stereo analysis to generate candidates, i.e., combining two different sensor types, like VS and TIR to perform stereo [60] or a VS stereo pair matched with TIR imaging [57]. This approach, which corresponds to sensor fusion (detailed in Section 2.7), is worth mentioning here because of its potential to widen the range of working conditions, e.g., in the case of the tetra-camera configuration, consisting of a VS pair for daytime and a TIR pair for nighttime.

Motion-based. Interframe motion and optical flow [61] have been used for foreground segmentation, primarily in the general context of moving obstacle detection. Franke and Heinrich [62] proposed to merge stereoprocessing, which extract depth information without time correlation, and motion analysis, which is able to detect small gray value changes in order to permit early detection of moving objects. In [63], Leibe et al. present a real-time Structure-from-Motion-based approach for ground plane estimation. This online estimated plane is used to update the camera calibration, and thus, to segment objects from the ground surface.

2.2.2 Analysis

The exhaustive scan is typically used in general human detection systems, e.g., image retrieval, whereas PPSs tend to use some kind of segmentation, as shown in Table 2. In fact, the latter can take advantage of some application prior knowledge (e.g., it is not necessary to search the top area of the image) so that the number of ROIs to process can be greatly reduced. For example, a typical exhaustive scan on a 640×480 image can provide from 200,000 to 1,000,000 ROIs, depending on the sampling step and the minimum ROI size. In contrast, sampling just the estimated road can reduce this number to 20,000-40,000, again depending on the density of the scan. Furthermore, stereo-based segmentation could further reduce this number by at least a factor of 10, depending on the content of the scene.

According to the literature, stereo is the most successful option. Two-dimensional-based analysis does not provide convincing results at this stage. For instance, symmetry is not very reliable, so extra cues such as depth are necessary, hot spot analysis seems to be ruled by heuristics, and attentional bottom-up pixel-based algorithms like [36] do not provide accurate ROI positions, so the reduction of the number of candidates is not as large as expected. More sophisticated appearance-based techniques are likely to be used during classification, not during candidates generation. In addition, the accuracy of motion-based approaches depends on driving speeds, and the reliability of those approaches has not been demonstrated under the wide range of ADAS conditions.

Stereo-based systems present several advantages: 1) They have good accuracy in the working range of pedestrian detection, 2) they are robust to circumstantial variability (e.g., illumination in VS or temperature in TIR), and 3) they provide useful information for other modules (e.g., distance estimations for tracking) and other ADAS applications (e.g., free space analysis [54], [64]). The drawbacks of such systems are as follows: 1) blind areas in nontextured regions, 2) slow speed (although advances in parallel data processing are being studied [65]), and 3) a requirement for postprocessing in order to separate regions with similar disparity and position to fit the pedestrian size and aspect ratio [58].

In conclusion, stereo is the primary option for future systems. Along with the aforementioned properties, stereo pairs are improving in accuracy, computation time, and resolution, facilitating the development of new systems. From our review, we conclude that an exhaustive study of the effect of the baseline of stereo pairs on pedestrian detection accuracy in urban environments is necessary. To the best of our knowledge, only [26] and [56] specify baseline information: 12 and 30 cm, respectively. A study of how baseline and depth accuracy parameters affect PPSs would implicitly relate the maximum vehicle speed and pedestrian distance.

It is clear that in such an oriented application problem, the use of scene prior knowledge plays a key role. A few recent studies on more sophisticated algorithms based on preattentive cues and context should be noted: Torralba and coauthors [66], [67] and Hoiem et al. [23], [68] groups. In these papers, the important roles of perspective, scene object dependencies, surfaces, and occlusions in object detection are demonstrated. In addition, active sensors (e.g., laser scanners), which can estimate distances without high computation times (Section 2.7), are also likely to be exploited in specific PPS tasks (e.g., short-time collision detection).

2.3 Object Classification

The object classification module receives a list of ROIs that are likely to contain a pedestrian. In this stage, they are classified as pedestrian or nonpedestrian aiming, with the goal of minimizing the number of false positives and false negatives.

2.3.1 Review

The approaches to object classification are purely 2D, and can be broadly divided into silhouette matching and appearance.

Silhouette matching. The simplest approach is the binary shape model, presented by Broggi et al. [39], in which an upper body shape is matched to an edge modulus image by simple correlation after symmetry-based segmentation. A more sophisticated approach is the Chamfer System, a silhouette-matching algorithm proposed by Gavrila et al. [55], [69], [70]. This system consists of a hierarchical template-based classifier (Fig. 3) that matches distance-transformed ROIs with template shapes in a coarse-to-fine manner. The shape hierarchy is generated offline by a clustering algorithm. This technique has also been exploited for TIR images in [51]. Also in the TIR spectrum, Nanda and Davis [71] perform probabilistic template matching on a multiscale basis by using just three templates (each for a defined scale). In [44], Broggi et al. present two methods that rely on templates, one of which is based on simple matching and another based on leg position.

Appearance. The methods included in this group define a space of image features (also known as descriptors), and a

TABLE 1
Different Learning Algorithms Used in PPS

Learning Alg.	Definition	Properties	Used Features
SVM [72]	Finds a decision boundary by maximizing the margin between the different classes.	- Decision boundary can be linear. - Data can be of any type, i.e., scalar or vector features, intensity images, etc.	Intensity image [46], [73] Haar Wav. [32], [74], [75], HOG [31], [43], [76], [77], Edgelet [76], Shape Cont. [77].
AdaBoost [78] ([80], [81], [82], [83])	Constructs a *strong* classifier by attaching *weak* classifiers (often rule-of-thumb) in an iterative greedy manner. Each new classifier focuses on missclassified instances.	- Speed optimized thanks to the use of cascades. - Can be combined with any classifier to find weak rules (e.g., with SVM [33]) - Few parameters to tune.	Haar [79], [30], [51], HOG [33], [77], EOH [30], Edgelet [76], [84], Shape Context. [77]
Neural Networks [86]	Different layers of neurons (many different configurations are possible) provide a non-linear decision.	- Many configurations and parameters to choose. - Raw data is often used, i.e., no explicit feature extraction process is needed.	Intensity image [85], Gradient mag. [56], [58], LRF [87].

classifier is trained by using ROIs known to contain examples (pedestrians) and counterexamples (nonpedestrians). Table 1 summarizes the typical learning algorithms (classifiers) used in the literature.

Following a holistic approach (i.e., target is detected as a whole) in [55], [70], Gavrila et al. propose a classifier that uses image gray-scale pixels as features and a *neural network with local receptive fields* (NN-LRFs [87]) as the learning machine that classifies the candidate ROIs generated by the Chamfer System. In [58], Zhao and Thorpe use image gradient magnitude and a *feedforward neural network*.

Papageorgiou and Poggio [32] introduce the so-called Haar wavelets (HWs) as features to train a quadratic support vector machines (SVMs) with front- and rear-viewed pedestrians. HWs compute the pixel difference between two rectangular areas in different configurations (Figs. 4a, 4b, and 4c), which can be seen as a derivative at a large scale. Viola and Jones [79], [88] propose AdaBoost cascades (layers of threshold-rule weak classifiers) as a learning algorithm to exploit Haar-like features (the original HWs plus two similar features, Figs. 4d and 4e), for surveillance-oriented pedestrian detection. In this case, HWs are also exploited to model motion information. These features have been quite successful for object recognition. Mählisch et al. [51] combined Haar-like features with the Chamfer System in a system based on TIR imagery. Recently, Gerónimo et al. [30] made use of Real AdaBoost to select the best features among a set of Haar-like and *edge orientation histograms* (EOHs; [89]) features to classify ROIs in the VS. EOHs first compute the gradient magnitude of the image, and then, distribute the pixels into k different bins (in this case, $k = 4$) according to their gradient orientation. The features are defined as the ratio between the summed gradient magnitudes of two bins for a given rectangular region. For example, the feature $\frac{\psi_{(0,\pi/4)}}{\psi_{(\pi,3\pi/4)}}$, where $\psi_{a,b}$ corresponds to the summed gradient magnitude of pixels laying in the specified (a, b) angle interval, gives one real value, which is fed as a feature to the threshold rule classifier. Both Haar-like features and EOH can make use of the *integral image representation* [79], which computes the sum of pixels of a region in just four memory accesses.

Dalal and Triggs [31] present a human classification scheme that uses SIFT-inspired [90] features, called *histograms of oriented gradients* (HOGs), and a linear SVM as a learning method. An HOG feature also divides the region into k orientation bins (in this case, $k = 9$), but instead of computing the ratio between two bins, they define four different cells that divide the rectangular feature, as illustrated in Fig. 7. In addition, a Gaussian mask is applied to the magnitude values in order to more heavily weight the center pixels, and the pixels are interpolated with respect to pixel location within a block (both factors disallow the use of the integral image). The resulting feature is a 36D vector containing the summed magnitude of each pixel cells, divided into 9 bins. These features have been extensively exploited in the literature. In [43], they are used for ADAS-oriented pedestrian detection in TIR images, while Dalal et al. [91] use them with optical flow images. Other papers propose new learning approaches, making use of the same features. Zhu et al. [33] use HOG as a weak rule of AdaBoost, achieving the same detection performance, but with less computation time, whereas Pang et al. [92] use Multiple Instance Learning (concretely, *Logistic Multiple Instance Boost* [83]) together with weak classifiers based on graph embedding to model variations in pedestrians' poses and viewpoints. Recently, Maji et al. [93] have outperformed the state-of-the-art detectors by using multilevel edge energy features (similar to HOG, but simpler) and the Intersection Kernel SVM. First, they apply nonmaximum suppression to each gradient orientation bin, which is then used to construct a pyramid of histogram features at different scales (64×64, 32×32, 16×16, 8×8). Intersection kernels are then used to train these features on an SVM.

Wu and Nevatia [94] studied the performance of short segments (up to 12 pixels long) of lines or curves, referred to as *edgelets*, as features for AdaBoost for VS images. In this case, a mask is attached to each feature, in order to provide pixelwise segmentation (Fig. 6). The same authors study edgelets and HOG together with AdaBoost and SVM learning algorithms, in both the VS and TIR [76]. Also, exploiting local gradient orientation features, Sabzmeydani and Mori [95] propose to use AdaBoost to model each $n \times n$ cell (they test different cell sizes, $n = 5, 10, 15$), with respect

TABLE 2
Visible Spectrum

Authors	Foreground Segmentation	Object Classification	Verification/Refinement	Tracking	Sensor
Gavrila et al. (1999–2006) [55], [69], [70]	Stereo+PSC	Hierarchical Template Matching (Chamfer System [69]) NN-RBF based on texture [87]	Stereo	$\alpha - \beta$ tracker [70] Silhouette, texture, stereo, fed to particle filters [117]	stereo
Bertozzi et al. (2003–04) [40], [41]	Symmetry+PSC		Stereo, PSC, ad hoc image filters, 3D curves matching	Kalman, grey-level, stereo	stereo
Broggi et al. (2000–03) [39], [38]	Stereo (v-disparity), PSC, Symmetry, head and shoulders model matching.		Stereo, PSC, entropy	—	stereo
Shashua et al. (2004) [97]	Texture (not detailed) +PSC	Parts-based gradient orientation and magnitude + Ridge Regression and AdaBoost	Multiframe after tracking: gait pattern, inward motion analysis scores (coupled with egomotion), goodness of classification	(used but not detailed)	mono
Grubb et al. (2004) [56]	Stereo (v-disparity)	Gradient magnitude+ Quadratic SVM	Classification goodness over time with help of tracking	Kalman+stereo	stereo
Zhao et al. (2000) [58]	Stereo+PSC	Gradient magnitude+ Neural Networks	—	—	stereo
Soga et al. (2005) [59]	Stereo+PSC	Four Directional Features (FDF) +Gaussian Kernel SVM	—	Extended Kalman Filter	stereo
Gerónimo et al. (2006-2007) [30]	Stereo-based Horizon Line Estimation [26] + PSC	Haar Wavelets and Edge Orientation Histograms + Real AdaBoost	—	—	stereo
Parra et al. (2005-2007) [98]	Stereo matching + lanes and road surface filtering + subtractive clustering	Parts-based SVM using edges and graylevel orientation, magnitude and texture	—	Kalman	stereo
Leibe et al. (2005) [63]	Scale-invariant Top-down Implicit Shape Model [127], [128]		Combination of Chamfer matching and its overlap with the segmentation	—	mono
Elzein et al. (2003) [61]	Interframe motion and optical flow	Euclidean distance of Haar Wavelets between templates and ROI	—	—	mono
Papageorgiou et al. (1997–2001) [32], [74]	—	- Holistic: Haar W.+ Q.SVM [32] - Parts-based: Haar W.+ Q/L.SVM [74]	—	Heuristic integration through time	mono
Dalal/Zhu et al. (2005/06) [31], [33]	—	Histograms of Oriented Grad. + - Linear SVM [31] - Linear SVM and AdaBoost cascade [33]	—	—	mono
Wu et al. (2007) [84]	—	Part-based multi-view edgels + Nested Weak Classifiers AdaBoost [82] Bayes-based parts combination		Parts+color+ appearance-based Mean-shift	mono

to its orientation. Each of the selected cells is referred to as a *shapelet* feature.

Tuzel et al. [96] propose a novel algorithm that is based on the covariance of different measures (position, first and second-order derivatives, gradient module, and gradient orientation) in subwindows as features, along with LogitBoost [81] using Riemannian manifolds. The achieved performance is comparable to state-of-the-art detectors, while the computation time is comparable to [31].

Other features and learning algorithms used in the literature include the gradient magnitude and quadratic SVM [56], Four Directional Features and Gaussian kernel SVM [59], and intensity image with Convolutional Neural Networks [85] or with an SVM [46], [73].

TABLE 3
Infrared Spectrum

Authors	Foreground Segmentation	Object Classification	Verification/Refinement	Tracking	Sensor
Broggi et al. (2006) [44]	Double threshold and binarization	Geometrical moments (eccentricity, object and legs inclination) model matching using probability	—	—	mono (NIR)
Sun et al. (2004) [138]	Threshold	Polar coordinates shape + SVM based on gray-level image features + parts-based classif.	—	—	mono (NIR)
Andreone et al. (2005) [75]	PSC	Heuristic pixel-based discarding process + SVM using Haar wavelets	—	—	mono (NIR)
Fang et al. (2004) [47]	Vertical projection Object contours	Histogram-, inertia-, contrast-based matching versus one unique pedestrian template	—	—	mono (TIR)
Bertozzi et al. (2003–05) [42], [20], [48] [116]	Hot areas, greyscale/edges vertical symmetry + Ad hoc filters + PSC [20], [42]		2D [20] / 3D [42], [113] model matching	Kalman filter with past history analysis [116]	mono (TIR)
	Horizontal/vertical histogram + Stereo [48]				stereo (TIR)
Suard et al. (2006) [43]	Simple threshold and bounding box generation	Histograms of Oriented Gradients + Linear SVM	—	—	mono (TIR)
Mählisch et al. (2005) [51]	PSC + Hypermutation Network (pixel classification [49])	Fusion of Chamfer Contour Matching and a Haar W.–based classifier	Multiple filter approach based on area over-lapping of windows	Second order motion model with Particle filter for state estimation	mono (TIR)
Xu et al. (2005) [73]	Threshold on histogram–equalized image + PSC + ground extraction over edges map	Intensity image + SVM (head/entire body based)	—	Kalman, Mean-Shift Method	mono (TIR)
Tsuji et al. (2002) [45]	Threshold: intermediate value of brightness distribution histogram	—	—	Stereo triangulation + egomotion using yaw sensor	stereo (TIR)

Part-based approaches, contrary to the previous holistic techniques, combine the classification of different parts of the pedestrian body (e.g., head and legs), instead of classifying the entire candidate as a single entity.

Mohan et al. [74] use HWs and a quadratic SVM to independently classify four human parts (head, legs, right arm, and left arm). The classifications of these parts are combined with a linear SVM. In [97], Shashua et al. use 13 overlapping parts (Fig. 5), described by SIFT inspired [90] features, and ridge regression to learn the classifier of each part. The training set is divided into nine clusters according to pose and illumination conditions, resulting in $9 \times 13 = 117$ classifiers, in order to deal with high intraclass variability. The outputs of the classifiers are fed as weak rules to an AdaBoost machine that sets the final classification rule. Wu and Nevatia [84] propose the use of four body parts (full body, head-shoulder, torso, and legs) and three view categories (front/rear, left profile, and right profile) to train a nested-weak-classifier AdaBoost [82]. They use edgelets as features. In this case, Bayesian reasoning, together with a typical surveillance assumption (camera looking down the plane), is used to combine the body parts. In the case of [98], Parra et al. define the features as the cooccurrence matrix between Canny edges and normalized gray-scale image, the orientation histogram, the magnitude and orientation of the image gradient, and the texture unit number, which are then fed to an SVM classifier. Tran and Forsyth [99] propose the estimation of the pedestrian pose in the ROI by the use of structure learning, which provides a tree parts configuration. After the estimation, the ROI conditioned by this configuration can be classified.

Felzenszwalb et al. [100] sum the classification score of the ROI and six different dynamic parts (they are not constrained to a unique position relative to the ROI). In this case, the authors use what they call *latent SVM* and HOG. Dollár et al. [101] extend the aforementioned Multiple Instance Learning to a part-based scheme called Multiple Component Learning, using Haar features. Here, gradient magnitude and orientation features are used. Notably, both approaches [100], [101] avoid the task of manually annotating parts since they are automatically determined by the method.

Lin and Davis [102] have recently proposed a technique that combines some of the aforementioned paradigms to a greater or lesser extent, i.e., silhouette, appearance, holistic,

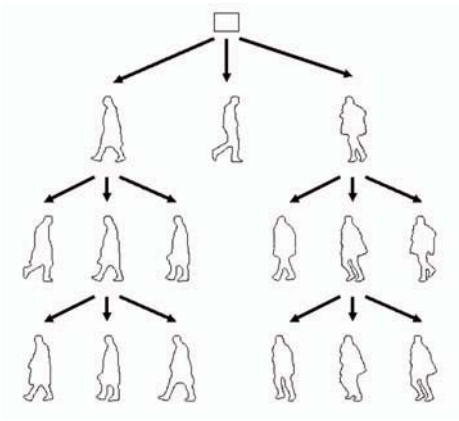

Fig. 3. Hierarchy of templates used in the Chamfer System by Gavrila (figure from [69]).

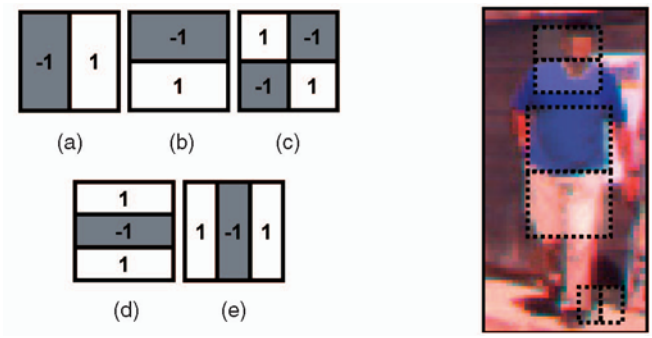

Fig. 4. (a), (b), (c), (d), and (e) Haar wavelets and Haar-like features, applied at specific positions of a pedestrian sample [32], [51], [30], [74].

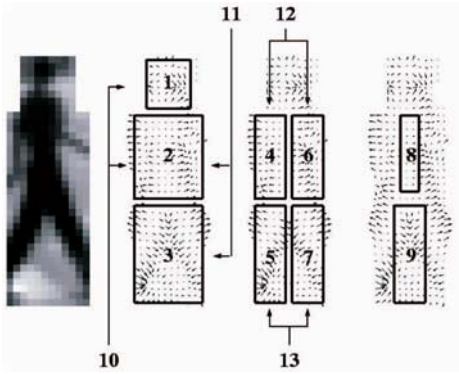

Fig. 5. Part-based classification using gradient-based features by Shashua et al. (figure from [97]).

and parts-based (see Fig. 8). First, HOG descriptors are computed for the whole image following Dalal's method. Then, the descriptors are used to extract a silhouette, which is fed to a probabilistic hierarchical part-matching algorithm. Finally, HOGs are again computed for the closest regions of the matched silhouette, serving as features for a radial basis function (RBF) kernel SVM.

Other approaches. Following recent research in object detection, Leibe et al. [103] present a technique termed as the *implicit shape model*, which avoids the ROI generation step. The idea is to use a keypoints detector, Hessian-Laplace [104] in this case, then compute a shape context descriptor [105] for each keypoint, and finally, cluster them to construct a codebook. During recognition, each detected

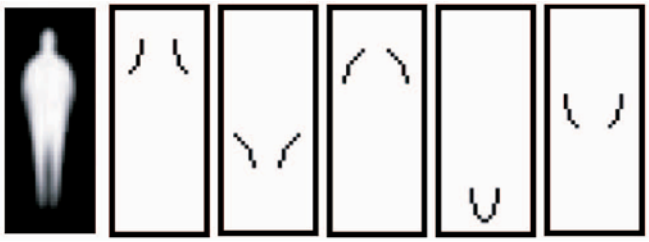

Fig. 6. First five edgelet features selected by AdaBoost in the approach by Wu and Nevatia (figure from [94]).

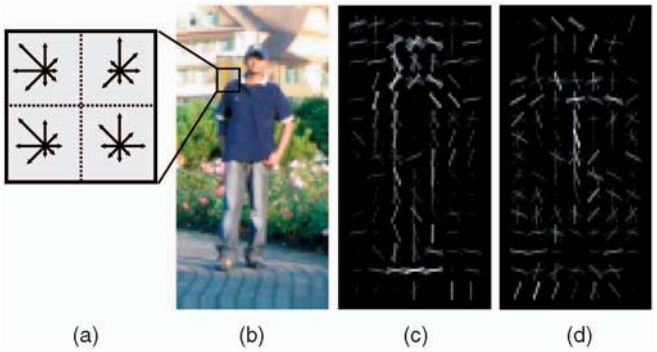

Fig. 7. Histograms of oriented gradients by Dalal and Triggs (figure from [31]). (a) The descriptor block. (b) Block placed on a sample image. (c) and (d) HOG descriptor weighted by positive and negative SVM weights.

Fig. 8. Pose invariant algorithm by Lin and Davis (figure from [102]). (a) Input image. (b) Part-template detections. (c) Pose and shape segmentation. (d) Cells grid used for HOG computation. (e) HOGs. (f) Cells relevant to HOG.

keypoint is matched to a cluster, which then votes for an object hypothesis using Hough voting, thus avoiding a candidate generation step. The Chamfer distance is used to provide a fine silhouette segmentation of the pedestrian. In [106], [107], Seeman et al. improve this technique with multiaspect (viewpoint and articulation) detection capabilities, extending the hypothesis voting to object shapes, rather than just objects.

2.3.2 Analysis

Silhouette matching methods are not applicable as standalone techniques. Even the elaborate Chamfer System needs an extra appearance-based step. In contrast, methods that exploit appearance seem to indicate the current direction of research, specifically revolving around the continuous development of new learning algorithms and features for use in this algorithms, not only in pedestrian detection but also in general object classification.

Despite the large number of papers, approaches tend to be poorly compared to one another in PPS research. Wojek et al.

[34] try shed some light on the comparison of classifiers with a study on some popular features and learning methods. Two conclusions are highlighted: HOGs and shape context features are the best option, independent of the learning algorithm, and feature combination significantly improves detector performance. In recent years, however, the lack of comparisons has been amended due to Dalal's proposal (both detector [31] and database [31]), which has been established as a *de facto* baseline. In fact, many of the techniques proposed within the last two years [96], [99], [95], [93], [100], [101], [102] use this benchmark, which makes it feasible to gain insights into the proposed module.

Given the number of papers presented in recent years, it is not possible to point to one method as the best option. Nevertheless, some research directions are clearly gaining relevance. Holistic classifiers seem to have reached their performance limit, at least for current databases, and are unable to deal with high variability. According to experiments, nonstandard poses greatly affect their performance: Beyond the usual straight versus crossing legs, pose variability also affected the head and torso alignment in the training examples. In addition, the diversity of poses causes many pedestrians to be poorly represented during training (e.g., running people, children, etc.). Parts-based algorithms that rely on dynamic part detection [100], [99], [107] handle pose changes better than holistic approaches. This information has been demonstrated to be beneficial in classification. Furthermore, other interesting ways to overcome this variability are being explored (e.g., multiple instance learning), which may provide additional benefits, like relaxing the annotation process. Of course, any improvement in existing algorithms, like the proposals in [93], [107], or new features that exploit typical measures (i.e., intensity, gradient, etc.) in new ways, like shape context [105] or HOG [31], will contribute to the improvement of these systems.

The real-time requirements of PPSs have been a principal restriction on the features and algorithms; however, high computational cost is not necessarily a handicap. For instance, there are works focused on optimizing algorithms, like [35], which can compute HOGs roughly seven times faster than was shown in the original proposal. In this case, a multiresolution rejection scheme is used. Additionally, computational power increases yearly, and hardware implementations of existing algorithms are being proposed. For example, two different GPU versions of the HOG [108], [34] work 10 and 34 times faster than the original one, respectively.

2.4 Verification/Refinement

Many systems contain one step that verifies and refines the ROIs classified as pedestrians. The verification step filters false positives, using criteria that do not overlap with the classifier, while the refinement step performs a fine segmentation of the pedestrian (not necessarily silhouette-oriented) to provide an accurate distance estimation or to support the subsequent tracking module.

2.4.1 Review

Gavrila et al. [55], [70] verify detections by performing cross correlation between the left image of a stereo pair and the isolated silhouette computed by the Chamfer system in the right image. In [53], Franke and Gavrila suggest the analysis of gait pattern of pedestrians crossing perpendicular to the camera. The target must be tracked before applying this method; thus, the order of verification/refinement and tracking modules is interchanged for this particular technique. Chamfer matching is used to both verify and refine the found pedestrian shapes in [109].

In [97], Shashua et al. propose a multiframe approval process that consists of validating the pedestrian-classified ROIs by collecting information from several frames: gait pattern, inward motion, confidence of the single-frame classification, etc. In this case, verification follows tracking.

For refinement, one essential algorithm that provides one detection per target is nonmaximum suppression. Assuming that classifiers provide a peak at the correct position and scale of the target and weaker responses around it, Dalal [110] makes use of mean shift [111] to find the minimum set of ROIs that best adjust to the pedestrians in the image. For the sake of completeness, it is worth mentioning the work by Agarwal et al. [112]. Their proposal, which was tested for vehicle detection instead of pedestrian detection, consists of two algorithms. The first creates an activation map, where high-confidence detections mark their neighborhoods as invalid for new detections. Given that this system is based on a parts-based classifier, the second algorithm constrains the parts to be assigned to only one detection, and thus, nonmaximum detections are discarded by iteratively decreasing their confidence.

In [39], by Broggi et al., the silhouette of the head and shoulders that is matched during classification is taken as a reference for refining detection down to the feet by using the vertical edges computed for the symmetry detection. The accurate location of the feet is then used to compute the distance to pedestrians by assuming a planar road. Then, stereoprocessing completes the refinement, by correlating the left-image bounding box to certain positions of the right image.

Some techniques that utilize TIR images are 2D model matching [20], 3D [42], [113] model matching, symmetry [114], and a multiple filter approach, based on the area overlap between positively classified ROIs and group multiple detections in a single window [51].

2.4.2 Analysis

This module should be a complement to the classification module. In fact, authors often refer to the described techniques as a two detection process, in the sense that verification algorithms are tied to the classification output, i.e., to the characteristics of its false positives. For instance, a classifier that fails to discard trees will not gain much benefit from a verifier that just distinguishes vertical regions in 3D. It is important to note that stereoinformation tends to be used as long as the classification is based on a 2D image. In addition, it is reasonable to expect that as more cues are used in verification, the results will be richer, e.g., stereoimaging may be combined with classification confidence, symmetry, or gait. It should also be noted that the use of verification after tracking presents an interesting approach since, potentially, common movement-based techniques (e.g., gait pattern analysis) used in surveillance

can be applied. The disadvantage of this approach is that this procedure is limited to walking pedestrians with clearly visible legs. This restriction usually implies that the pedestrian is close to the camera, which means that the latency of the analysis is a very important issue.

The employed refinement methods are chosen according to the utilized foreground segmentation technique and the available information. Each of the methods presents advantages and disadvantages. For instance, mean shift has proven to be a reliable technique for full-scan processing, but has not been evaluated for other foreground segmentation algorithms for which the ROI scan is not very dense (e.g., road sampling or stereo). Distance estimations from stereoimages, when available, are a good cue for adjustment of the final ROI size, but the error increases with the target distance. A study of the quality of the final detections in terms of road plane adjustment (i.e., pedestrian distance), bounding box accuracy, etc., under a set of different candidate generation schemes, refinement algorithms, and cues (e.g., disparity, road plane adjustment, and TIR symmetry) would be of great interest.

2.5 Tracking

The most evolved systems use a tracking module to follow detected pedestrians over time. This step has several purposes: avoiding false detections over time, predicting future pedestrians positions, thus feeding the foreground segmentation algorithm with precandidates, and, at a higher level, making useful inferences about pedestrian behavior (e.g., walking direction).

2.5.1 Review

Franke et al. propose the use of two Kalman filters [54], [115], one controlling lateral motion (yaw rate of the own vehicle is used) and the other controlling longitudinal motion, to determine the speed and acceleration of detected objects. Later, in [55], [70], authors from the same research group used an α-β tracker (a simplified Kalman filter with preestimated steady-state gains and a constant velocity model) based on the bounding box representation from their stereo verification phase. Three cues were used: the euclidean distance between bounding box centroids, shape dissimilarities (to avoid multiple tracks for single objects), and the Chamfer distance (to avoid multiple objects assigned to single tracks). Also, using Kalman filters as tracking filters, Bertozzi et al. [41] use ROI overlapping to merge tracks, Binelli et al. [116] enrich the predictions with egomotion computed from velocity and yaw sensors, and Grubb et al. [56], in addition to Kalman filtering, use Bayesian probability to provide certainty, trajectory, and speed of pedestrians over time.

Particle filters are also widely used in tracking. Giebel et al. [117] use them to track multiple objects in 3D (in this case, the cues are silhouette, texture, and stereo). Philomin et al. [118] use the Condensation method [119] (a variant of particle filters) to track silhouettes approximated by B-Splines. Arndt et al. [120] employ particle filters in a track-before-detect paradigm by coupling the tracking algorithm to a cascade classifier [121]. The real-time GPU implementation of particle filters by Mateo and Otsuka [122] is also worth mentioning.

Recently, Leibe et al. [63] proposed the use of a color model and what they refer to as the *event cone*, i.e., the space-time volume in which the trajectory of a tracked object is sought. The authors claim that, although this proposal relies on the same equations as the Kalman filter, it is superior to it in the sense that object state estimation can be based on several previous steps and multiple trajectories for the observed data can be evaluated.

Zhang et al. [123] propose the use of network flows to optimize association of detections to tracks. A min-cost flow algorithm is used to perform the detection-track association, and an explicit occlusion model is used to control long-term occlusions.

Research on detection in crowded scenarios has recently led to coupled detection-tracking frameworks which share information between both modules instead of treating them as independent stages. Gammeter et al. [124] perform multibody tracking by combining the *implicit shape model* detector [103] and the stereo-odometry-based tracker of [29]. Each trajectory is passed to a single-person articulated tracker, which estimates the 3D pose and dynamics of each individual. Adnriluka et al. [125] detect targets using a part-based detector, and then use a Gaussian process latent variable model to compute the temporal consistency of detections over time. Finally, Singh et al. [126] use the output from the part-based detector in [84] to initialize tracklets (short track segments of high confidence detections) and residuals (low confidence detections). The tracklet descriptors (based on color, motion, and 3D height) and tracklet paths (using multiple hypotheses) are then associated within a global optimization framework.

2.5.2 Analysis

Tracking represents an important aspect of transforming a pedestrian detection algorithm into a PPS. However, this module has not received as much attention as other modules; each paper presents its own proposal and no comparisons have been made. Hence, it is not easy to extract conclusions. It can be said that the Kalman filter is by far the most heavily utilized algorithm, but tracking cues range from simple 2D ROI localization to color, silhouette, texture, or 3D information. Recent coupled detection-tracking algorithms represent a promising way to exploit richer tracking cues, e.g., tracking independently detected body parts instead of complete, rigid pedestrian silhouette models.

In our opinion, although there are many interesting proposals, much work remains to be done in tracking benchmarking (Section 3) before solid conclusions can be reached on this topic.

2.6 Application

The last module of a PPS takes high-level decisions based on the information from previous modules. This module represents a complete area of research, which includes psychological issues, human-machine interactions, and many issues that are out of the scope of this survey. The reader is referred to [129], [45], [130] for a description of application modules in PPSs.

2.7 Sensors and Fusion

As previously stated, all of the reviewed techniques rely on the output of cameras. They are the most widely used

TABLE 4
Sensor Fusion

Authors	Foreground Segmentation	Object Classification	Verification/Refinement	Tracking	Sensor
Fardi et al. (2005) [133]	Laserscanner map + Active contour on TIR to extract shape	Euclidean distance of Fourier descriptors between object and reference sets (on TIR)	Walking cycle analysis using optical flow and egomotion sensor	Kalman filter + data fusion	TIR (mono) + laser
SAVE-U (2005) [129]	Radar + Vision-based Local Histograms on Edge Images (visible and TIR) and PSC	NN-RBF [87]	Fusion of radar and tracking info	α–β tracker	TIR + visible (mono) + radar
Milch et al. (2001) [135]	Radar + velocity and steering sensors	Active contours (TIR or visible) using a trained shape model	—	—	TIR or visible (mono) + radar
Bertozzi et al. (2006-2007) [137], [114], [139]	v-disparity + disparity image + hotspots (TIR)	Symmetry + template matching + vertical edge-based refinement		—	TIR + visible (both stereo)
Premebida et al. (2007) [134]	Laserscanner-based tracking of points	GMM on laserscanner data + AdaBoost (Haar Wav.) + Bayesian fusion of Classifiers	—	Kalman on laserscanner performed as foreground segmentation	Visible (mono) + laserscanner

sensors, due to the high potential of visual features, high spatial resolution, and richness of texture and color cues. However, it was clear throughout the review that image analysis is far from simple: cluttering and illumination, among many other factors, affect the performance. Furthermore, the VS can be affected by glaring sources of light, while TIR can be influenced by other *hot* objects (e.g., engines of other vehicles or light poles), changing weather conditions (i.e., relative temperature changes), year/season, etc., [131]. In fact, pedestrians could be warmer or colder than the background, depending on such factors [132].

Fusion of VS/TIR sensors and active sensors, which are used to obtain complementary information, is being investigated in the context of on-board pedestrian detection. The strengths and weaknesses of different kinds of sensors can be complemented in order to improve the overall system performance. Active sensors are based on technologies that emit signals and observe their reflection from the objects in the environment, for example, radars emitting radio waves or laser scanners emitting infrared light. In general, these sensors are convenient for detecting objects and providing superior range estimates out to larger distances relative to passive sensors.

Next, we review some systems that implement sensor fusion. Table 4 provides a summary of the most relevant systems.

2.7.1 Review

Fardi et al. [133] combine a laser scanner with a TIR shape extraction method to select ROIs, using Kalman filtering as the data fusion algorithm. Premebida et al. [134] segment and track clusters of points along the 1D laser scanner dimension (note that tracking and segmentation are performed together), while classification is performed using data from the laser scanner (using a Gaussian mixture to model clusters centroid, standard deviation, radius, etc.) and the VS (using Haar wavelets and AdaBoost). Milch and Behrens [135] make use of radar, velocity, and steering sensors to generate hypotheses. They then perform classification using a shape model for either the VS or the TIR spectrum images. Linzmeier et al. [136] also exploit radar, but combine it with thermopile, steering angle, and ambient temperature sensors. In this case, fusion can be done at a low level (ROI generation combining radar and thermopile) and a high level (ROIs independently generated by all the sensors).

Combining VS and TIR spectra from the two camera types has also been proposed. In [137], by Bertozzi et al., v-disparity is computed using VS, and then, foreground segmentation is carried out in both the VS and TIR (2D area overlapping and 3D information are the fusion cues). Finally, symmetry and template matching are used to classify, verify, and refine final detections in the TIR. Krotosky and Trivedi [57] evaluate tetra and tri-sensor systems that utilize both the VS and TIR. For example, in the tri-sensor approach, a VS stereo pair performs ROI generation, while VS, TIR, and disparity-based HOG-like features are fed to an SVM for classification.

One of the systems of the SAVE-U project [129] attaches a radar sensor to the VS and TIR cameras. They implement three different levels of fusion: sensor, low, and high level. The first level is aimed at associating different radar detections (each from an independent sensor) into unique real objects, as well as establishing a correspondence between the VS and TIR images. For low-level fusion, ROIs are first detected in the VS by an algorithm based on histogram edge orientations, then resized by radar data (i.e., they are adjusted to the ground by using the accurate radar distance estimation), and finally, classified by NN-LRF (Section 2.3). High-level fusion associates the radar and VS information (distance and azimuth) of each object, and then, tracks their trajectories.

2.7.2 Analysis

Sensor fusion for ADAS is an open area of research, and much work is still required before convincing results will be achieved in real scenarios. The ideal combination of sensors must be clarified, given that each sensor has its own failure cases. For example, the conclusions of the SAVE-U project [129] state that, although the radar-camera combination worked well in simple test tracks, the radar became unreliable at 10-15 m when working in real scenes due to reflections from other objects (humans have low reflectance). Laser scanners, which work with infrared beams, are progressively gaining the interest of researchers as they can detect pedestrians while providing accurate distance estimates. However, laser scanners are affected by adverse weather conditions just like cameras, which is not the case for radar.

3 Benchmarking

In contrast to other areas, like face detection or document analysis, pedestrian detection for ADAS lacks well-established databases and benchmarking protocols. The absence of realistic public databases and the difficulty of implementing published techniques have usually led researchers to evaluate new proposals with local private databases, without any comparison to other state-of-the-art proposals. Public databases are necessary for two reasons: 1) to evaluate algorithms with different example sets, taken at different places under different conditions, but specifically from different research groups (which adds extra variability) and 2) to compare new algorithms with existing ones, i.e., given that it is hard to reproduce algorithms, the easiest way of establishing comparisons is to compare results from the same databases following the same criteria.

3.1 Classification Benchmarking

At the moment, only two pedestrian databases specifically built for ADAS are publicly available: the Daimler Chrysler (DC) Pedestrian Classification Benchmark [87] and the Computer Vision Center (CVC) Pedestrian Database [30]. They contain samples taken from moving vehicles in urban scenarios, viewed from the same road plane (i.e., no large camera tilts as in surveillance), under the typical ADAS situations. DC features very low-resolution samples (18×36 pixels), while CVC samples maintain their original image size (from 140×280 to 12×24 pixels).

Other, non-ADAS person databases can also be used as long as the appearance of people is relevant to ADAS (i.e., seen from the same plane, standing, etc.). In this case, three popular databases can be used: the MIT Pedestrian Data set [32], which is almost perfectly classified in [31], and thus, is outdated; the INRIA Person Data set [31], which is currently quite popular for general human classification evaluation, but contains a large number of samples taken from high-resolution photographs; and the USC Pedestrian Detection Test Set [84], which contains ADAS-like pedestrian samples, divided into front/rear full view, front/rear partial interhuman occlusions, and front-/rear-/side-viewed pedestrians.

Table 5 summarizes these databases, and Fig. 9 illustrates some positive samples. Note that the number of samples refers to the annotated real pedestrian samples; this number is often increased by mirroring or pixel-shifting the window (e.g., in the DC database, where the final number of samples is 24,000).

TABLE 5
Public Pedestrian Databases

Name	Samples	Number	Range/Views
DC[3][87]	18×36 greyscale	4,000 pos. 25,000 neg.	Far (~30-50m) F/S/B
CVC[4][30]	Orig. size color	1,000 pos. 6,175 neg.	All range (5-50m) F/S/B
MIT[5][32]	64×128 color	924 pos. no neg.	Near F/B
INRIA[6][31]	Orig. size & 64×128 color	1,239 pos. 1,218 neg. im.	Near F/S/B
USC[7][84]	Orig. size grayscale	816 pos. no negs	Varied F/B

Views legend: F (front), S (side), and B (back).
[3] *http://www.science.uva.nl/research/isla/downloads/pedestrians.*
[4] *http://www.cvc.uab.es/adas/databases.*
[5] *http://cbcl.mit.edu/software-datasets.*
[6] *http://lear.inrialpes.fr/data.*
[7] *http://iris.usc.edu/Vision-Users/OldUsers/bowu/DatasetWebpage/dataset.html.*

3.2 Evaluation Protocols

The simplest protocol for evaluating classifier performance is to classify a set of samples other than the training ones, called the testing set, and plot performance curves, e.g., receiver operating characteristic (ROC), detection-error trade-off (DET), etc. We call this protocol the *database-based test*, and is used in [31], [87], [30]. Another approach is to test the classifier on full images, for instance, by classifying all of the possible ROIs, and then plotting the same curves (e.g., precision-recall) according to some criterion that decides whether a positive detection is correct or not, e.g., more than 50 percent overlap with an annotated pedestrian. This approach is used in [110] and in the well-known PASCAL Challenge;[8] we call it the *full-image-based test*. The first protocol is used to evaluate the classifier module, while the second is more convenient for evaluating the overall system results. Other performance measures consist of segmentation side accuracy and segmentation side efficiency, which are used to evaluate detection in the TIR by using a full-image-based approach [47], and trajectory/alarms, which take into account the tracking and application modules [70].

Table 6 summarizes data on the performance of a few systems. Columns 2-5 correspond to the parameters and performance of the utilized classifiers, while 6-8 refer to complete system results when available. Note that the number of negative samples in every classifier varies; thus, the rates of the false positives cannot be directly compared.

3.3 Future Needs

Although these two protocols are helpful as standard benchmarking tools, there are still some measures that require definition. For instance, there are no standard methods for measuring cases of overlapping detections or spurious false detections or for testing different methods of

8. http://pascallin.ecs.soton.ac.uk/challenges/VOC.

Fig. 9. Some annotated pedestrian samples in two PPS databases (DC and CVC) and a general human detection database (INRIA). Image quality and resolution of INRIA samples are clearly superior to samples in the other two databases.

TABLE 6
Systems Performance

Authors	Learning ROI Size	Classifier Training Set	Classifier Testing Set	Classifier Performance	System Testing Set	System Performance	Range, speed
Gavrila et al. [55], [69] [69], [70]	(Chamfer) 70–102 pixels wide (NN-LRF) 18×36	(Chamfer) 1,250 pos (NN-LRF) 14 400 pos 15 000 neg	(Chamfer) 900 images (NN-LRF) 9 600 pos 10 000 neg	(Chamfer) 60–90% DR n/a FPR (NN-LRF) 90% DR 10% FPR	24min driving	(all) 52–76% DR 30% precision (risky) 80–90% DR 75% precision	$5-25m$ $4-10Hz$
Shashua et al. [97]	12×36	25 000 pos 25 000 neg	15 244 in total	93.5% DR 8% FPR	5hr driving	(inward moving) 96%DR, 1 FP (statio. inpath) 93%DR, 3 FP (statio. outpath) 85%DR, 102 FP	$3-25m$ $20-25Hz$
Grubb et al. [56]	–	1 500 pos 20 000 neg	150 pos 2 000 neg	75% DR 2% FPR	2 500frame (14 pedest.)	83.5% DR 0.4% FPR	till $30m$ $23Hz$
Zhao et al. [58]	30×65	1 012 pos 4 306 neg	254 pos 363 neg	85.4% DR 0.05 % FPR	FGS ROIs	85.2% DR 3.1% FPR	$3-12Hz$
Soga et al. [59]	52 pix high	14 052 pos 30 841 neg	1 200 pos 25 601 neg	90% DR 0.3% FPR	4 sequences	83.3% DR avg 0.007 avg/frame	till $40m$ $10fps$
Fang et al. [47]	n/a	n/a	n/a	n/a	3 sequences of 40 secs. each	(winter) $\sim$ 100%DR at $\sim$ 0 FPR (summer1) 85% DR 10% FPR (summer2) 78% DR 10% FPR	$6fps$

DR stands for detection rate, FPR stands for false positive rate. All of the systems use their own private benchmark databases.

classifier ROI selection (i.e., full scan [31], PSC-based, or more complex techniques [97]).

There is still room for improvement in the databases, specifically with regard to the following aspects:

- Quantity. New improvements in face detection can be evaluated using more than 20 databases, whereas PPS can be tested on only five.
- Representivity. For instance, databases do not usually contain images of children or very tall people, thus leaving them out of typical PSC. However, it is clear that children crossing the road (e.g., running behind a ball) present a very plausible scenario.
- Variability. In addition to clothes, pose, and illumination, it would be interesting to increase the variability in terms of height, distance, degree of occlusion, and complexity of the background (here, virtual pedestrian synthesis [140] could be very helpful).
- Resolution. Some databases [31] contain well-focused images of pedestrians from a photographic camera, which usually corresponds to targets close to the vehicle in a PPS. Hence, it is difficult to determine both the working distance of a given classifier and whether the results can be extrapolated to detect further targets, which tend to be blurred and contained within a small number of pixels.
- Sensors. Until now, we have referred to visible spectrum data sets, but if we look for a public TIR ADAS-oriented database, we will not find one. To

the best of our knowledge, the OTCBVS[9] [141] is the only publicly available TIR database, but it is not suited to ADAS benchmarking. The same issue presents itself with active sensors, such as laser scanner or radar.

Future challenges of PPSs not only require new public databases, but complete annotated sequences, i.e., common benchmarking protocols and databases must not be constrained to classification, but also extended to other modules and whole systems. In this sense, we list some guidelines for future database development that would be of great interest to the community: full train and test sequences (i.e., not only still images) for evaluating whole systems; foreground segmentation ground truth (e.g., road area or vertical candidate regions in 3D) would lead to specific studies on this module; tracking ground truth is useful for evaluating the proposed algorithms and features for tracking; and especially, richer pedestrian annotation, in terms of view or distance, to train and evaluate new object classification paradigms, such as multiclass algorithms.

Two new databases that may address some of these challenges are expected to be presented soon, according to personal communication with the authors [142], [143].

4 Discussion

A perfect on-board PPS must detect the presence of people in the way of the vehicle and react according to the risk of running over the pedestrian (warn the driver, brake the vehicle, deploy external airbags, and perform an evasive maneuver), without disturbing the driver if there is no risk at all. Moreover, such a system should work well independent of the time, road, and weather condition. Additionally, the cost of the pedestrian detection module should be relatively small compared to the total cost of the vehicle.

It is clear in the reviewed literature that, in the last decade, an enormous research effort has been made in automatic people detection. The scientific community has made significant advances; however, at the same time, we are not even halfway to an ideal PPS.

In order to illustrate this situation, we follow a line of reasoning inspired by [9]. Let us take the classifier proposed by Shashua et al. [97], which achieves a 95 percent detection rate at a 10 percent false positive (FP) rate under realistic PPS conditions (i.e., hard negative samples, not over the whole image, and the smallest ROIs to classify are 12×36 pixels). If we have 200,000 ROIs generated by an exhaustive scan, this classifier will provide 20,000 FP/frame. If the number of candidates is reduced to 1,000 per image by using some foreground segmentation technique, the number of FP is reduced to 100. Moreover, the authors claim to reduce this number to only 75 ROIs that need to be checked. This number represents 7.5 FP per image, which corresponds to 187.5 FP per second at 25 fps. A tracking module could filter out detections to reduce the final number to approximately 1 FP/s; however, even 1 FP/s (i.e., 60 FP/minute) is not sufficient for a PPS.

In addition, we note that the work done up to now in pedestrian detection covers only a subset of the possible challenges. A few relevant questions that have to be addressed are provided below.

- It is assumed that the full body of the pedestrian is visible, even in part-based methods, given that they need a reliable classification of parts. Handling occlusions is a relevant issue (e.g., pedestrians can suddenly appear from behind a parked vehicle) and it is not clear how such methods work if the pedestrian is only partially seen.
- Nighttime pedestrian detection has only barely been addressed. The limited studies that have addressed it did so in the context of the TIR spectrum. Importantly, drivers need more help at nighttime (poor illumination, contrast, color, etc.).

Once the main limitations of current proposals are pointed out, we would like to share our point of view regarding how to address such a challenge from the computer vision side.

4.1 First Intermediate Useful Challenge

The pursuit of a perfect PPS must be considered a long-term goal. The development of a PPS that works under restricted conditions is already useful. For instance, a system that works only in daytime, under good weather conditions (no heavy rain/snow/fog), over a range of distances up to 50 m is, from our viewpoint, the first intermediate challenge for the community. According to [144], these conditions represent a very relevant scenario in accidents.

4.2 Imaging Technology

According to the reviewed literature, the most promising option is pedestrian detection based on stereo rigs with HDR cameras, given the ability of modern stereo techniques to provide useful 3D information out to about 50 m and the ability of HDR to provide well-contrasted NIR images.

Although TIR images have been quite popular during recent years for nighttime scenarios, it is unclear whether this technology will ultimately be chosen for serial production. Unfortunately, such cameras are quite expensive, have low resolution, are more difficult to integrate since they cannot see through windshields, and are not producing quite convincing results. In our opinion, although TIR-based research is interesting, we think that NIR sensors deserve more attention, given the following points: First, modern headlight systems cover the visible near infrared ranges of the spectrum and they have new motion capabilities, thus providing better visibility than standard low beams [145]. Second, NIR images are quite similar to VS images, and so daytime systems can be more easily adapted to NIR imagery than to TIR imagery. Third, there are other ADAS applications like *lane departure warning* or *traffic sign recognition* for which TIR imagery does not seem appropriate, in contrast to NIR. Therefore, in order to cover such applications along pedestrian detection, both spectra are needed even if we rely on TIR for the latter task.

4.3 Improving Overall System Performance

We refer to both real-time reactions as well as to detection rate versus false alarms. Although it is clear that any improvement in the individual modules should result in better system performance, one can see that most of the

9. http://www.cse.ohio-state.edu/otcbvs-bench.

research has been focused on the object classification task, specifically focused on detection performance. Researchers have tried to improve features and learning machines by increasing their discrimination power or lowering their computation time. Going forward with this research is fully justified. We would like to offer a few suggestions:

- Multiclass approaches should be incorporated, not only to consider different pedestrian models, but also to check for other targets (e.g., vehicles) and increase the robustness of the system.
- It may be interesting to see if 3D measures can be used with 2D information to improve classifiers.
- Given that pedestrians closer to the camera are seen with more detail than ones farther away, a study on the benefits of training different classifier models depending on the target distance would be of interest.
- Finally, research into part-based methods, in addition to the benefits when dealing with pose variability, must also focus on partially occluded pedestrians. This amendment would result in PPSs with lower detection latency in some critical cases.

Despite the potential improvements in detection, it is unrealistic to think of a perfect classification module. After all, the goal is to have neither misdetections nor false alarms at the system level. Therefore, we should think of how all modules can contribute to these goals.

The high relevance of foreground segmentation must be stressed, either as an isolated module or integrated with the classification module. The benefits are twofold: 1) Given that classification tends to be the most time-consuming task, reduction of the image area to be processed also reduces the overall system time; 2) by submitting fewer background regions for classification, the rate of false alarms can also be reduced while the same detection rate is maintained. In addition, this module can also be useful for selecting complex negative examples, instead of the typical random ones. Therefore, the learning machine could concentrate on discriminating complex foreground from pedestrians while avoiding background, and the associated performance curves would be more significant and realistic. Finally, we must point out that, although 3D information is our preferred approach, fusion with 2D preattentive cues and context can conveniently produce gains in robustness.

The need to deal with egomotion in outdoor scenarios and with a changing background is probably the reason why tracking for PPS has not received as much attention as in other applications, like surveillance. However, this module has considerable potential to improve the overall system performance. Final classification based on several frames is more robust than classification based on a single frame since additional features can be collected [79] along the temporal axis and temporal coherence analysis can be performed, etc. As an example, since distant pedestrians appear smaller in the image, they tend to be more difficult to classify; hence, a track-before-detect strategy can be reliable. For closer targets, the latency of the system must be lower, so a detect-before-track strategy is expected. Fortunately, closer targets present more detail, and classification is easier. In addition, future work should make use of the aforementioned information obtained from the vehicle (e.g., speed and yaw rate) and image stabilization obtained by several techniques [24], [26].

Throughout this review, we have shown that active sensors (e.g., radar and laser scanner) are a good solution to the problem of obtaining 3D information in real time, i.e., for performing foreground segmentation. As an example, a laser scanner with an HDR camera seems to be a straightforward option for performing ROI generation and classification. The challenge for the computer vision community is to develop a system that is able to beat active sensors-based setups, especially since computer vision techniques are cheaper to maintain than active sensors. For instance, a setup using a stereo rig based on HDR cameras could provide reliable foreground segmentation and classification during both the day and night.

In addition to the aforementioned ways of improving PPSs, a complementary strategy actually under investigation [146] is the concept of *driver in the loop*, i.e., taking into account the *driver state*. Since the aim of a PPS is to assist the driver, not to substitute for them, it is not necessary to bother the driver with information if they are clearly paying attention to the road. On the contrary, the PPS can warn the driver about risky pedestrians on road areas that they are not monitoring (e.g., pedestrians suddenly appearing from a lateral direction). However, more research into driver monitoring (e.g., developing databases of synchronized driver and outdoor images) and psychological aspects (i.e., the danger of drivers intentionally paying less attention due to the PPS) is required.

5 Conclusion

Intelligent vehicles represent a key technology for reducing the number of accidents between pedestrians and vehicles. Given the difficulties that such systems must overcome, i.e., real-time detection of changing targets in uncontrolled outdoor scenarios, pedestrian protection is by no means an easy task. Consequently, a plethora of research papers addressing the challenge have been presented during the last decade. We have reached a point where a state-of-the-art review can be of great help to summarize all of the work done to this point.

In this paper, we have presented a survey that does not follow the typical one-by-one paper review. We think that following such a strategy would obfuscate the similarities and differences between proposals when addressing specific aspects of the challenge. Instead, we have cast the pedestrian detection challenge as a task composed of subtasks, and proposed an architecture of logic modules, each with attached responsibilities. In our opinion, such a reviewing strategy more clearly presents the state of the art from the point of view of the community, especially work from novel researchers in the area. Accordingly, we have reviewed and analyzed the literature while trying to explicitly elucidate how the different subtasks are addressed in each work. Specifically, 108 papers published between 1996 and 2008 (16 from 2008) have been reviewed. Moreover, we have included a discussion section, where we presented our viewpoint of the pedestrian detection field, given the reviewed papers as well as our own experience, and pointed out current weaknesses and possible future trends.

The primary conclusion is that in the last decade, there has been an enormous research effort in automatic people detection. However, the feeling is that we are still far from developing an ideal system. It is clear that major progress has been made in pedestrian classification, mainly due to synergy with generic object detection and applications such

as face detection and surveillance. However, there is still work to do before a useful performance level is achieved and protection systems can be installed in a serial car. We have argued that such a responsibility must be shared between tasks like foreground segmentation and tracking. We have emphasized a short-term need for realistic and public databases in order to standardize performance evaluation. Altogether we are optimistic about future achievements in this field of research.

ACKNOWLEDGMENTS

The authors would like to thank Professor Bernt Schiele and the anonymous referees whose invaluable comments significantly improved the paper. They also thank D.M. Gavrila, A. Shashua, Y. Gdalyahu, N. Dalal, Z. Lin, and B. Wu for giving permission to reprint their figures. This work was supported by the Spanish Ministry of Education and Science under projects TRA2007-62526/AUT and Consolider Ingenio 2010: MIPRCV (CSD200700018). David Gerónimo was supported by the Spanish Ministry of Education and Science and European Social Fund grant BES-2005-8864.

REFERENCES

[1] D. Gavrila, P. Marchal, and M.-M. Meinecke, "SAVE-U, Deliverable 1-A: Vulnerable Road User Scenario Analysis," technical report, Information Soc. Technology Programme of the EU, 2003.
[2] W. Jones, "Building Safer Cars," *IEEE Spectrum,* vol. 39, no. 1, pp. 82-85, Jan. 2002.
[3] United Nations—Economic Commission for Europe "Statistics of Road Traffic Accidents in Europe and North America," 2005.
[4] R. Bishop, *Intelligent Vehicle Technologies and Trends.* Artech House, Inc., 2005.
[5] L. Vlacic, M. Parent, and F. Harashima, *Intelligent Vehicle Technologies.* Butterworth-Heinemann, 2001.
[6] T. Heinrich, "Bewertung Von Technischen Maßnahmen zum Fußgängerschutz am Kraftfahrzeug," technical report, Technische Univ. Berlin, 2003.
[7] T. Moeslund, A. Hilton, and V. Krüger, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis," *Computer Vision and Image Understanding,* vol. 104, nos. 2/3, pp. 90-126, 2006.
[8] D. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan, *Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis.* Now publishers, 2005.
[9] D. Gavrila, "Sensor-Based Pedestrian Protection," *IEEE Intelligent Systems,* vol. 16, no. 6, pp. 77-81, Nov./Dec. 2001.
[10] T. Gandhi and M.M. Trivedi, "Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 976-981, 2006.
[11] T. Gandhi and M.M. Trived, "Pedestrian Protection Systems: Issues, Survey, and Challenges," *IEEE Trans. Intelligent Transportation Systems,* vol. 8, no. 3, pp. 413-430, Sept. 2007.
[12] M.-H. Yang, D.J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 1, pp. 34-58, Jan. 2002.
[13] Z. Sun, G. Bebis, and R. Miller, "On-Road Vehicle Detection: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 5, pp. 694-711, May 2006.
[14] D. Gerónimo, A. López, and A. Sappa, "Computer Vision Approaches for Pedestrian Detection: Visible Spectrum Survey," *Proc. Third Iberian Conf. Pattern Recognition and Image Analysis,* pp. 547-554, 2007.
[15] S. Nayar and V. Branzoi, "Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time," *Proc. Int'l Conf. Computer Vision,* vol. 2, pp. 1168-1175, 2003.
[16] S. Marsi, G. Impoco, A. Ukovich, S. Carrato, and G. Ramponi, "Video Enhancement and Dynamic Range Control of HDR Sequences for Automotive Applications," *EURASIP J. Advances in Signal Processing,* vol. 2007, p. 9, 2007.
[17] P. Knoll, "HDR Vision for Driver Assistance," *High-Dynamic-Range (HDR) Vision,* B. Hoefflinger, ed., pp. 123-136, Springer, 2007.
[18] A. Broggi, M. Bertozzi, and A. Fascioli, "Self-Calibration of a Stereo Vision System for Automotive Applications," *Proc. IEEE Int'l Conf. Robotics and Automation,* pp. 3698-3703, 2001.
[19] T. Dang and C. Hoffmann, "Stereo Calibration in Vehicles," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 268-273, 2004.
[20] M. Bertozzi, A. Broggi, M. Carletti, A. Fascioli, T. Graf, P. Grisleri, and M.-M. Meinecke, "IR Pedestrian Detection for Advanced Driver Assistance Systems," *Proc. Pattern Recognition Symp.,* pp. 582-590, 2003.
[21] A. Broggi, P. Grisleri, T. Graf, and M.-M. Meinecke, "A Software Video Stabilization System for Automotive Oriented Applications," *Proc. Vehicular Technology Conf.,* vol. 5, pp. 2760-2764, 2005.
[22] L. Bombini, P. Cerri, P. Grisleri, S. Scaffardi, and P. Zani, "An Evaluation of Monocular Image Stabilization Algorithms for Automotive Applications," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 1562-1567, 2006.
[23] D. Hoiem, A. Efros, and M. Hebert, "Putting Objects in Perspective," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 2137-2144, 2006.
[24] R. Labayrade, D. Aubert, and J. Tarel, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry through 'v-Disparity' Representation," *Proc. IEEE Intelligent Vehicles Symp.,* vol. 2, pp. 17-21, 2002.
[25] Z. Hu and K. Uchimura, "U-V-Disparity: An Efficient Algorithm for Stereovision Based Scene Analysis," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 48-54, 2005.
[26] A. Sappa, F. Dornaika, D. Ponsa, D. Gerónimo, and A. López, "An Efficient Approach to Onboard Stereo Vision System Pose Estimation," *IEEE Trans. Intelligent Transportation Systems,* vol. 9, no. 3, pp. 476-490, Sept. 2008.
[27] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, T. Graf, and R. Schmidt, "High Accuracy Stereovision Approach for Obstacle Detection on Non-Planar Roads," *Proc. IEEE Intelligent Eng. Systems,* pp. 211-216, 2004.
[28] A. Ess, B. Leibe, and L. VanGool, "Depth and Appearance for Mobile Scene Analysis," *Proc. Int'l Conf. Computer Vision,* 2007.
[29] A. Ess, B. Leibe, K. Schindler, and L. VanGool, "A Mobile Vision System for Robust Multi-Person Tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[30] D. Gerónimo, A. Sappa, A. López, and D. Ponsa, "Adaptive Image Sampling and Windows Classification for On-Board Pedestrian Detection," *Proc. Fifth Int'l Conf. Computer Vision Systems,* 2007.
[31] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 886-893, 2005.
[32] C. Papageorgiou and T. Poggio, "A Trainable System for Object Detection," *Int'l J. Computer Vision,* vol. 38, no. 1, pp. 15-33, 2000.
[33] Q. Zhu, S. Avidan, M.-C. Yeh, and K.-T. Cheng, "Fast Human Detection Using a Cascade of Histrograms of Oriented Gradients," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 1491-1498, 2006.
[34] C. Wojek, G. Dorkó, A. Schulz, and B. Schiele, "Sliding-Windows for Rapid Object Class Localization: A Parallel Technique," *Proc. Symp. German Assoc. for Pattern Recognition,* pp. 71-81, 2008.
[35] W. Zhang, G. Zelinsky, and D. Samaras, "Real-Time Accurate Object Detection Using Multiple Resolutions," *Proc. Int'l Conf. Computer Vision,* pp. 1-8, 2007.
[36] F. Miau, C. Papageorgiou, and L. Itti, "Neuromorphic Algorithms for Computer Vision and Attention," *Proc. Int'l Symp. Optical Science and Technology,* pp. 12-23, 2001.
[37] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
[38] A. Broggi, A. Fascioli, I. Fedriga, A. Tibaldi, and M.D. Rose, "Stereo-Based Preprocessing for Human Shape Localization in Unstructured Environments," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 410-415, 2003.
[39] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-Based Pedestrian Detection," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 215-220, 2000.

[40] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape-Based Pedestrian Detection and Localization," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 328-333, 2003.

[41] M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi, R. Chapuis, and F. Chausse, "Pedestrian Localization and Tracking System with Kalman Filtering," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 584-589, 2004.

[42] M. Bertozzi, A. Broggi, A. Fascioli, T. Graf, and M.-M. Meinecke, "Pedestrian Detection for Driver Assistance Using Multiresolution Infrared Vision," *IEEE Trans. Vehicular Technology,* vol. 53, no. 6, pp. 1666-1678, Nov. 2004.

[43] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian Detection Using Infrared Images and Histograms of Oriented Gradients," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 206-212, 2006.

[44] A. Broggi, R. Fedriga, A. Tagliati, T. Graf, and M.-M. Meinecke, "Pedestrian Detection on a Moving Vehicle: An Investigation about Near Infra-Red Images," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 431-436, 2006.

[45] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of Night-Vision System," *IEEE Trans. Intelligent Transportation Systems,* vol. 3, no. 3, pp. 203-209, Sept. 2002.

[46] Q. Tian, H. Sun, Y. Luo, and D. Hu, "Nighttime Pedestrian Detection with a Normal Camera Using SVM Classifier," *Proc. Int'l Symp. Neural Networks,* pp. 189-194, 2005.

[47] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki, "A Shape-Independent Method for Pedestrian Detection with Far-Infrared Images," *IEEE Trans. Vehicular Technology,* vol. 53, no. 6, pp. 1679-1697, Nov. 2004.

[48] M. Bertozzi, A. Broggi, A. Lasagni, and M.D. Rose, "Infrared Stereo Vision-Based Pedestrian Detection," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 24-29, 2005.

[49] M. Oberländer, "Hypermutation Networks—A Discrete Approach to Machine Perception," *Proc. Third Weightless Neural Networks Workshop,* 2005.

[50] U. Meis, M. Oberländer, and W. Ritter, "Reinforcing the Reliability of Pedestrian Detection in Far-Infrared Sensing," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 779-783, 2004.

[51] M. Mählisch, M. Oberländer, O. Löhlein, D. Gavrila, and W. Ritter, "A Multiple Detector Approach to Low-Resolution FIR Pedestrian Recognition," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 325-330, 2005.

[52] U. Franke and I. Kutzbach, "Fast Stereo Based Object Detection for Stop & Go Traffic," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 339-344, 1996.

[53] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler, "Autonomous Driving Goes Downtown," *IEEE Intelligent Systems,* vol. 13, no. 6, pp. 40-48, Nov./Dec. 1999.

[54] U. Franke and A. Joos, "Real-Time Stereo Vision for Urban Traffic Scene Understanding," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 273-278, 2000.

[55] D. Gavrila, J. Giebel, and S. Munder, "Vision-Based Pedestrian Detection: The PROTECTOR System," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 13-18, 2004.

[56] G. Grubb, A. Zelinsky, L. Nilsson, and M. Rilbe, "3D Vision Sensing for Improved Pedestrian Safety," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 19-24, 2004.

[57] S. Krotosky and M.M. Trivedi, "On Color-, Infrared-, and Multimodal-Stereo Approaches to Pedestrian Detection," *IEEE Trans. Intelligent Transportation Systems,* vol. 8, no. 4, pp. 619-629, Dec. 2007.

[58] L. Zhao and C. Thorpe, "Stereo and Neural Network-Based Pedestrian Detection," *IEEE Trans. Intelligent Transportation Systems,* vol. 1, no. 3, pp. 148-154, Sept. 2000.

[59] M. Soga, T. Kato, M. Ohta, and Y. Ninomiya, "Pedestrian Detection with Stereo Vision," *Proc. IEEE Int'l Conf. Data Eng. Workshops,* p. 1200, 2005.

[60] S. Krotosky and M.M. Trivedi, "Multimodal Stereo Image Registration for Pedestrian Detection," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 109-114, 2007.

[61] H. Elzein, S. Lakshmanan, and P. Watta, "A Motion and Shape-Based Pedestrian Detection Algorithm," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 500-504, 2003.

[62] U. Franke and S. Heinrich, "Fast Obstacle Detection for Urban Traffic Situations," *IEEE Trans. Intelligent Transportation Systems,* vol. 3, no. 3, pp. 173-181, Sept. 2002.

[63] B. Leibe, N. Cornelis, K. Cornelis, and L. VanGool, "Dynamic 3D Scene Analysis from a Moving Vehicle," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[64] H. Badino, W. Franke, and R. Mester, "Free Space Computation Using Stochastic Occupancy Grids and Dynamic Programming," *Proc. Int'l Conf. Computer Vision, Workshop Dynamical Vision,* 2007.

[65] W. vander Mark and D. Gavrila, "Real-Time Dense Stereo for Intelligent Vehicles," *IEEE Trans. Intelligent Transportation Systems,* vol. 7, no. 1, pp. 38-50, Mar. 2006.

[66] B. Hidalgo-Sotelo, A. Oliva, and A. Torralba, "Human Learning of Contextual Priors for Object Search," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 86-93, 2005.

[67] A. Torralba and A. Oliva, "The Role of Context in Object Recognition," *Trends in Cognitive Sciences,* vol. 11, no. 12, pp. 520-527, 2007.

[68] D. Hoiem, A. Efros, and M. Hebert, "Closing the Loop in Scene Interpretation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[69] D. Gavrila, "Pedestrian Detection from a Moving Vehicle," *Proc. European Conf. Computer Vision,* vol. 2, pp. 37-49, 2000.

[70] D. Gavrila and S. Munder, "Multi-Cue Pedestrian Detection and Tracking from a Moving Vehicle," *Int'l J. Computer Vision,* vol. 73, no. 1, pp. 41-59, 2007.

[71] H. Nanda and L. Davis, "Probabilistic Template Based Pedestrian Detection in Infrared Videos," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 15-20, 2002.

[72] V. Vapnik, *The Nature of Statistical Learning Theory.* Springer, 1995.

[73] F. Xu, X. Liu, and K. Fujimura, "Pedestrian Detection and Tracking with Night Vision," *IEEE Trans. Intelligent Transportation Systems,* vol. 6, no. 1, pp. 63-71, Mar. 2005.

[74] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 4, pp. 349-361, Apr. 2001.

[75] L. Andreone, F. Bellotti, A.D. Gloria, and R. Lauletta, "SVM-Based Pedestrian Recognition on Near-Infrared Images," *Proc. Fourth Int'l Symp. Image and Signal Processing and Analysis,* pp. 274-278, 2005.

[76] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian Detection in Infrared Images Based on Local Shape Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[77] C. Wojek and B. Schiele, "A Performance Evaluation of Single and Multi-Feature People Detection," *Proc. DAGM Symp.,* pp. 82-91, 2008.

[78] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Computer and System Sciences,* vol. 55, no. 1, pp. 119-139, 1997.

[79] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," *Proc. Int'l Conf. Computer Vision,* vol. 2, pp. 734-741, 2003.

[80] R. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-Rated Predictions," *Machine Learning,* vol. 37, no. 3, pp. 297-336, 1999.

[81] J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting," *Annals of Statistics,* vol. 28, no. 2, pp. 337-407, 2000.

[82] C. Huang, H. Ai, B. Wu, and S. Lao, "Boosting Nested Cascade Detector for Multi-View Face Detection," *Proc. Int'l Conf. Pattern Recognition,* vol. 2, pp. 415-418, 2004.

[83] X. Xu and E. Frank, "Logistic Regression and Boosting for Labeled Bags of Instances," *Proc. Pacific Asia Conf. Knowledge Discovery and Data Mining,* 2004.

[84] B. Wu and R. Nevatia, "Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Part Detectors," *Int'l J. Computer Vision,* vol. 75, no. 2, pp. 247-266, 2007.

[85] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian Detection with Convolutional Neural Networks," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 224-229, 2005.

[86] C. Bishop, *Neural Networks for Pattern Recognition.* Oxford Univ. Press, 1995.

[87] S. Munder and D. Gavrila, "An Experimental Study on Pedestrian Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 11, pp. 1863-1868, Nov. 2006.

[88] M. Jones and D. Snow, "Pedestrian Detection Using Boosted Features over Many Frames," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[89] K. Levi and Y. Weiss, "Learning Object Detection from a Small Number of Examples: The Importance of Good Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 53-60, 2004.
[90] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.
[91] N. Dalal, B. Triggs, and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance," *Proc. European Conf. Computer Vision,* pp. 428-441, 2006.
[92] J. Pang, Q. Huang, and S. Jiang, "Multiple Instance Boost Using Graph Embedding Based Decision Stump for Pedestrian Detection," *Proc. European Conf. Computer Vision,* vol. 4, pp. 541-552, 2008.
[93] S. Maji, A. Berg, and J. Malik, "Classification Using Intersection Kernel Support Vector Machines Is Efficient," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[94] B. Wu and R. Nevatia, "Simultaneous Object Detection and Segmentation by Boosting Local Shape Feature Based Classifier," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.
[95] P. Sabzmeydani and G. Mori, "Detecting Pedestrians by Learning Shapelet Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1-8, 2007.
[96] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian Detection via Classification on Riemannian Manifold," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 10, pp. 1713-1727, Oct. 2008.
[97] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian Detection for Driving Assistance Systems: Single-Frame Classification and System Level Performance," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 1-6, 2004.
[98] I. Parra, D. Fernández, M. Sotelo, L. Bergasa, P. Revenga, J. Nuevo, M. Ocana, and M.A. García, "Combination of Feature Extraction Method for SVM Pedestrian Detection," *IEEE Trans. Intelligent Transportation Systems,* vol. 8, no. 2, pp. 292-307, June 2007.
[99] D. Tran and D. Forsyth, "Configuration Estimates Improve Pedestrian Finding," *Proc. Conf. Neural Information Processing Systems Conf.,* pp. 1529-1536, 2007.
[100] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A Discriminatively Trained, Multiscale, Deformable Part Model," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[101] P. Dollár, B. Babenko, S. Belongie, P. Perona, and Z. Tu, "Multiple Component Learning for Object Detection," *Proc. European Conf. Computer Vision,* pp. 211-224, 2008.
[102] Z. Lin and L. Davis, "A Pose-Invariant Descriptor for Human Detection and Segmentation," *Proc. European Conf. Computer Vision,* vol. 4, pp. 423-436, 2008.
[103] B. Leibe, A. Leonardis, and B. Schiele, "Robust Object Detection with Interleaved Categorization and Segmentation," *Int'l J. Computer Vision,* vol. 77, nos. 1-3, pp. 259-289, 2008.
[104] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A Comparison of Affine Region Detectors," *Int'l J. Computer Vision,* vol. 65, nos. 1/2, pp. 43-72, 2005.
[105] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 4, pp. 509-522, Apr. 2002.
[106] E. Seeman, B. Leibe, and B. Schiele, "Multi-Aspect Detection of Articulated Objects," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 1582-1588, 2006.
[107] E. Seeman and B. Schiele, "Cross-Articulation Learning of Robust Detection of Pedestrians," *Proc. DAGM Symp.,* 2006.
[108] L. Zhang and R. Nevatia, "Efficient Scan-Window Based Object Detection Using GPGPU," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[109] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian Detection in Crowded Scenes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 878-885, 2005.
[110] N. Dalal, "Finding People in Images and Videos," PhD thesis, Inst. Nat'l Polytechnique de Grenoble/INRIA Rhône-Alpes, 2006.
[111] D. Comaniciu, "An Algorithm for Data-Driven Bandwidth Selection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 281-288, Feb. 2003.
[112] S. Agarwal, A. Awan, and D. Roth, "Learning to Detect Objects in Images via a Sparse, Part-Based Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 11, pp. 1475-1490, Nov. 2004.
[113] A. Broggi, A. Fascioli, P. Grisleri, T. Graf, and M.-M. Meinecke, "Model-Based Validation Approaches and Matching Techniques for Automotive Vision Based Pedestrian Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 3, p. 1, 2005.
[114] M. Bertozzi, A. Broggi, M. DelRose, and M. Felisa, "A Symmetry-Based Validator and Refinement System for Pedestrian Detection in Far Infrared Images," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 155-160, 2007.
[115] G. Welch and G. Bishop, "An Introduction to The Kalman Filter," technical report, Dept. of Computer Science, Univ. of North Carolina at Chapel Hill, 2002.
[116] E. Binelli, A. Broggi, A. Fascioli, S. Ghidoni, P. Grisleri, T. Graf, and M.-M. Meinecke, "A Modular Tracking System for Far Infrared Pedestrian Recognition," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 759-764, 2005.
[117] J. Giebel, D. Gavrila, and C. Schnör, "A Bayesian Framework for Multi-Cue 3D Object Tracking," *Proc. European Conf. Computer Vision,* pp. 241-252, 2004.
[118] V. Philomin, R. Duraiswami, and L. Davis, "Pedestrian Tracking from a Moving Vehicle," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 350-355, 2000.
[119] M. Isard and A. Blake, "Contour Tracking by Stochastic Propagation of Conditional Density," *Proc. European Conf. Computer Vision,* pp. 343-356, 1996.
[120] R. Arndt, R. Schweiger, W. Ritter, D. Paulus, and O. Löhlein, "Detection and Tracking of Multiple Pedestrians in Automotive Applications," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 13-18, 2007.
[121] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 511-518, 2001.
[122] O. Mateo and K. Otsuka, "Real-Time Visual Tracker by Stream Processing," *J. Signal Processing Systems,* 2008.
[123] L. Zhang, Y. Li, and R. Nevatia, "Global Data Association for Multi-Object Tracking Using Network Flows," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[124] S. Gammeter, A. Ess, T. Jäggli, K. Schindler, B. Leibe, and L. VanGool, "Articulated Multi-Body Tracking under Egomotion," *Proc. European Conf. Computer Vision,* 2008.
[125] M. Adnriluka, S. Roth, and B. Schiele, "People-Tracking-by-Detection and People-Detection-by-Tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[126] V. Singh, B. Wu, and R. Nevatia, "Pedestrian Tracking by Associating Tracklets Using Detection Residuals," *Proc. Workshop Motion and Video Computing,* pp. 1-8, 2008.
[127] B. Leibe, A. Leonardis, and B. Schiele, "Combined Object Categorization and Segmentation with an Implicit Shape Model," *Proc. European Conf. Computer Vision Workshop Statistical Learning in Computer Vision,* pp. 17-32, 2004.
[128] B. Leibe and B. Schiele, "Scale-Invariant Object Categorization Using a Scale-Adaptive Mean-Shift Search," *Proc. DAGM Symp.,* pp. 145-153, 2004.
[129] P. Marchal, M. Dehesa, D. Gavrila, M.-M. Meinecke, N. Skellern, and R. Viciguerra, "SAVE-U. Final Report," technical report, Information Soc. Technology Programme of the EU, 2005.
[130] T. Graf, K. Seifert, M.-M. Meinecke, and R. Schmidt, "Human Factors in Designing Advanced Night Vision Systems," *Proc. Fifth Congress and Exhibition on Intelligent Transport Systems and Services,* 2005.
[131] C.-Y. Chan and F. Bu, "Literature Review of Pedestrian Detection Technologies and Sensor Survey," technical report, Inst. of Transportation Studies, Univ. of California at Berkeley, 2005.
[132] E. Goubet, J. Katz, and F. Porikli, "Pedestrian Tracking Using Thermal Infrared Imaging," *Proc. SPIE Conf. Infrared Technology and Applications,* pp. 797-808, 2006.
[133] B. Fardi, U. Schuenert, and G. Wanielik, "Shape and Motion-Based Pedestrian Detection in Infrared Images: A Multi Sensor Approach," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 18-23, 2005.
[134] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A Lidar and Vision-Based Approach for Pedestrian and Vehicle Detection and Tracking," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems,* pp. 1044-1049, 2007.
[135] S. Milch and M. Behrens, "Pedestrian Detection with Radar and Computer Vision," *Proc. Conf. Progress in Automobile Lighting,* 2001.

[136] D. Linzmeier, M. Skutek, M. Mekhaiel, and K. Dietmayer, "A Pedestrian Detection System Based on Thermopile and Radar Sensor Data Fusion," *Proc. Int'l Conf. Information Fusion,* vol. 2, 2005.
[137] M. Bertozzi, A. Broggi, M. Felisa, G. Vezzoni, and M. DelRose, "Low-Level Pedestrian Detection by Means of Visible and Far Infra-Red Tetra-Vision," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 231-236, 2006.
[138] H. Sun, C. Hua, and Y. Luo, "A Multi-Stage Classifier Based Algorithm of Pedestrian Detection in Night with a Near Infrared Camera in a Moving Car," *Proc. Third Int'l Conf. Image and Graphics,* pp. 120-123, 2004.
[139] M. Bertozzi, A. Broggi, C. Hilario, R. Fedriga, G. Vezzoni, and M.D. Rose, "Pedestrian Detection in Far Infrared Images Based on the Use of Probabilistic Templates," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 327-332, 2007.
[140] M. Enzweiler and D. Gavrila, "A Mixed Generative-Discriminative Framework for Pedestrian Classification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.
[141] J. David and M. Keck, "A Two-Stage Approach to Person Detection in Thermal Imagery," *Proc. Workshop Applications of Computer Vision,* vol. 1, pp. 364-369, 2005.
[142] C. Wojek, S. Walk, and B. Schiele, "Multi-Cue Onboard Pedestrian Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2009.
[143] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: A Benchmark," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2009.
[144] A.-S. Karlsson et al., "Deliverable D50.30. User Needs State of the Art and Relevance for Accidents. PReVENT Project Apalaci: Preventive and Active Safety Applications," 2005.
[145] A. López, J. Hilgenstock, A. Busse, R. Baldrich, F. Lumbreras, and J. Serrat, "Nighttime Vehicle Detection for Intelligent Headlight Control," *Proc. Int'l Conf. Advanced Concepts for Intelligent Vision Systems,* pp. 113-124, 2008.
[146] S. Park and M. Trivedi, "Driver Activity Analysis for Intelligent Vehicles: Issues and Development Framework," *Proc. IEEE Intelligent Vehicles Symp.,* pp. 644-649, 2005.

David Gerónimo received the BSc degree in computer science from the Universitat Autònoma de Barcelona in 2004, and the MSc degree from the Computer Vision Center in 2006. He is currently working toward the PhD degree with the project "Pedestrian Detection in Advanced Driver Assistance Systems," granted by the Ministry of Science and Technology of Spain under the FPI grant. His research interests include stereo, feature selection, machine learning for candidate generation, and object detection.

Antonio M. López received the BSc degree in computer science from the Universitat Politécnica de Catalunya in 1992, the MSc degree in image processing and artificial intelligence from the Universitat Autònoma de Barcelona (UAB) in 1994, and the PhD degree in 2000. Since 1992, he has been giving lectures in the Computer Science Department of the UAB, where he currently is an associate professor. In 1996, he participated in the foundation of the Computer Vision Center at the UAB, where he has held different institutional responsibilities, presently being the responsible for the research group on advanced driver assistance systems by computer vision. He has been responsible for public and private projects, and is a coauthor of more than 50 papers, all in the field of computer vision.

Angel D. Sappa received the electromechanical engineering degree from the National University of La Pampa, General Pico, Argentina, in 1995, and the PhD degree in industrial engineering from the Polytechnic University of Catalonia, Barcelona, Spain, in 1999. In 2003, after holding research positions in France, the United Kingdom, and Greece, he joined the Computer Vision Center, where he is currently a senior researcher. He is a member of the Advanced Driver Assistance Systems Group. His research interests span a broad spectrum within the 2D and 3D image processing. His current research focuses on stereoimage processing and analysis, 3D modeling, and model-based segmentation. He is a member of the IEEE.

Thorsten Graf received the diploma (MSc) degree in computer science and the PhD degree (his thesis was on "Flexible Object Recognition Based on Invariant Theory and Agent Technology") from the University of Bielefeld, Germany, in 1997 and 2000, respectively. In 1997, he became a member of the "Task Oriented Communication" graduate program, University of Bielefeld, funded by the German research foundation DFG. In June 2001, he joined Volkswagen Group Research, Wolfsburg, Germany. Since then, he has worked on different projects in the area of driver assistance systems as a researcher and project leader. He is an author or coauthor of more than 15 publications and owns several patents. His research interests include image processing and analysis dedicated to advanced comfort/safety automotive applications.