Contents

Chapter 1
Introduction

Intelligent machines assist us in many situations of our everyday lives: medical diagnoses, communications, transportation, education, surveillance, etc. It is expected that during the current century machines will become even more integrated in our daily experiences, especially focused on improving aspects such as safety and comfort.

This book describes a type of systems aimed at avoiding pedestrian to vehicle collisions. These systems, formally known as Pedestrian Protection Systems (PPSs), must detect and track pedestrians, and provide the necessary outputs to the host vehicle in order to prevent potential accidents and even reduce their severity when unavoidable. The breakthrough of these systems occurred at the beginning of the twenty-first century thanks to the advances in sensing, the maturity of artificial intelligence and computer vision, and the increase in machines' computational power. In the last years intensive research efforts in this technology have been carried out by both public and private entities, which have projected their massive commercialization during this decade.

1.1 Automobile's Impact

When people are asked the most relevant technology that has changed most the landscapes of our cities during the last century, the automobile is a common answer. Indeed, the human development in the modern era is represented up to some extent by the automobile. It has changed societies not only in urban planning, industry and economy but also in demographic distribution and social interactions. Employment, leisure and relationships are all shaped to a greater or lesser degree by automobiles. This is clearly illustrated when comparing two families from the early 1900s and the 2000s. The former family bought their food and clothes in local stores, close to the living place, traveled to nearby working places, and the leisure travels were restricted to the nearby regions. Nowadays, most of the shopping activities are concentrated in

D. Gerónimo and A. M. López, *Vision-based Pedestrian Protection Systems for Intelligent Vehicles*, SpringerBriefs in Computer Science,
DOI: 10.1007/978-1-4614-7987-1_1, © The Author(s) 2014

large shopping malls out of the cities, working places are farther away and the leisure trips are not restricted to nearby regions but to whole countries and continents.

Like many other technologies, from the very beginning automobiles carried an undesirable dark side: traffic accidents. The first death by a motor vehicle was registered in Ireland on August 31st, 1869 [90]. Although the number of fatalities was low at the beginning, fatalities exponentially grew throughout the years given the popularization of cars. One and a half centuries later, road accidents represent the ninth cause of death worldwide, and the ONU predicts that in 2030 it will be the fifth [233]. Every year almost 1.2 million people are killed in traffic crashes, while the number of injuries rises to 50 million. Furthermore, attending to the increasing automobile productions in low and middle-income countries, these numbers are expected to rise considerably. Figure 1.1 illustrates the number of deaths per 100,000 population as a result of a traffic accident. As can be seen, high income countries and regions such as the United States or the European Union tend to have a lower number of fatalities than the rest, even though the number of vehicles in these countries is high. Before having a look at this map, one could have the wrong idea of thinking that a lower number of vehicles would be reflected in a lower number of traffic accidents and deaths. In the United States there are 0.8 vehicles per capita [66] (the USA is the most motorized country in the world). In Egypt this number is exactly the half [233], however the number of deaths per 100,000 inhabitants is three times the deaths in the USA. Nigeria, with only 0.3 vehicles per capita, has twice the number of deaths per 100,000 than the USA. In fact, it is the longstanding traffic regulation together with the consciousness-raising of this problem which have progressively decreased the number of fatal accidents in high-income countries. On the contrary, low-income

Fig. 1.1 Statistics of deaths related to traffic accidents in the world. It can be clearly seen that although rich countries have higher number of vehicles than low-income countries, the number of fatalities is lower thanks to the improved safety measures, government campaigns and regulation

countries tend to have a higher number of deaths given the opposite reason: lack of regulation in many aspects and low consciousness of the problem. Attending to this, as low-income countries evolve and start increasing their number of vehicles and transport networks the relevance of the problem arises.

According to the International Organization of Motor Vehicle Manufacturers [228], every year around 59 million passenger cars and 20 million commercial vehicles are produced worldwide. This represents an increase of 44 % and 17 % with respect to year 2,000 production, respectively. Although this rate will probably not sustain in very industrialized countries as a result of the increasing price of oil, the popularization of air travel and the *green* trends, it will be largely compensated by emerging economies such as China or India, with increases of 800 % and 400 % in total number of produced vehicles yearly from 2000 to 2010.

Figure 1.2 illustrates another dramatic fact related to the evolution of two emerging economies. On the one hand the United States and the European Union have average population and a big fleet of vehicles. On the other hand growing countries such as China or India have the biggest populations in the world but an average number of vehicles (between 0.2 and 0.5 per capita) [233]. As previously mentioned, having a smaller fleet does not lead to a lower number of accidents, as it is clearly appreciated in this figure. However, it is clear that as emerging countries increase their number of vehicles, the number of deaths will also rise if no solution is implemented. The most direct solutions to fix this problem are well-known given that they have been developed for decades: researching new technology to increase the safety of vehicles and infrastructures.

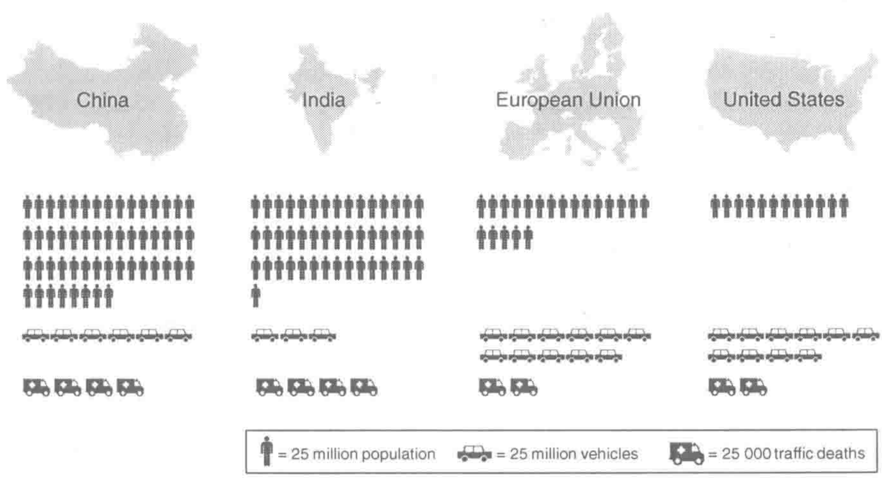

Fig. 1.2 Population, vehicle and traffic deaths statistics of two emerging economies (China and India) and two high-income ones (USA and European Union)

此为试读，需要完整PDF请访问：www.ertongbook.com

1.2 Advanced Driver Assistance Systems

From the beginning, the automobile industry has proposed new technologies aimed at improving safety and reduce the number of accidents. This technology has progressively gained complexity and improved performance through the years. The first electric headlamps were introduced in 1898 in the Columbia Electric Car, while turn signal lights were devised in 1907. In 1955 Ford included seat belts as an optional equipment and Saab made them standard in 1958, albeit the first proposals and tests of seat belts were made in the 1920s. Volvo presented the current standard three-points belt in 1959, which prevents the vehicle occupants to hit against the interior elements after an impact. The airbag was developed in the late 1960s, however, similarly as with the seat belts it took from two to three decades to become an standard in the United States and in Europe. Nowadays, it is estimated that around 12,000 lives are saved yearly by seat belts and 2,000 by airbags just in the United States [281]. Antilock braking system (ABS) and electronic stability control (ESC) were introduced by Bosch and Mercedes-Benz in 1978 and 1995, respectively. ABS is designed to avoid tires blocking in order to keep their grip constant. ESC is designed to avoid skidding by applying brakes to individual wheels. Both approaches were feasible thanks to the invention of electronics, which provided the necessary fast physics computations required. As can be seen, until the 1990s the technological advances in security relied mostly in physical devices focused on providing safety to the vehicle passengers when the accidents or dangerous situations were happening.

In the last twenty years, advanced driver assistance systems (ADAS) have aimed at predicting dangerous situations and anticipating accidents. These intelligent systems provide warnings and assist drivers to take decisions and even take automatic evasive actions in extreme cases. They are different from the previous technologies in the sense that they do not only rely on physical/mechanical cues from the host vehicle, but they can also *understand* the exterior world or the driver state. In this case, Artificial Intelligence plays a key role when pursuing this understanding. In addition, research in sensors, human machine interfaces and even psychology is also important for these systems.

The pioneer in ADAS was Ernst Dickmanns' group, from the Universitat der Bundeswehr München (Germany), who in 1986 presented an autonomous driving system able to drive through a controlled scenario (a closed highway) at up to 96 km/h [67]. The system consisted of cameras, rudimentary image processors and Kalman filtering. This research later lead to the first European project on autonomous vehicles, called Prometheus.

Nowadays many different ADAS may be found in the marked. For example, the adaptive cruise control (ACC) keeps a constant distance to the front vehicle by slowing down or accelerating the host vehicle. It was introduced by Mercedes and Jaguar in the late 1990s [157]. Lane departure warning (LDW) systems warn the driver when the vehicle moves out of its lane, unless the corresponding direction turn sign is on. It was introduced in trucks in 2000 and later in sedans [158]. This technology is being improved by assisting the steering action or warning/intervening in lane changing in case of danger. Finally, one of the currently hot research topics are advanced front lighting (AFL) systems, which control the headlight parameters so that the beam is optimized for different conditions such as driving speed and direction. The reader can refer to different publications that include comprehensive reviews on these systems [31, 311].

1.3 Pedestrian Protection Systems

Traditionally, the technological improvements in automobiles have been addressed to protect the vehicle occupants in vehicle-to-vehicle crashes. On the contrary, in terms of the automobile industry, road users such as pedestrians and bicyclists have not received the same attention. Having a look at the statistics it can be seen that the proportion of pedestrians killed in accidents is considerably high. In 2003 almost 150,000 injured and 7,000 killed pedestrians were reported in the European Union roads [78], representing the second source of injuries and fatalities just after four-wheeled vehicle passengers. The United States' numbers are similar, counting 70,000 injured and 4,000 killed [281]. In low and middle income countries this number can neither be neglected. As an example, just Delhi City (India) registered almost 1,000 killed pedestrians in 1994 [217], and at the moment this number is likely to have grown given that the number vehicles has been doubled in the country [280]. Statistics also state that 70 % of the people involved in car-to-pedestrian accidents were in front of the vehicle, of which 90 % were moving [137].

Figure 1.3 illustrates the percentage of pedestrian deaths over all the traffic deaths (vehicle passengers, motorbike passengers, cyclists, etc.). A relation can be seen, with some exceptions, between the percentage of pedestrian fatalities and the countries income. For example, in low income countries such as Dem. Rep. of Congo or Kenya more than 50 % of traffic deaths are pedestrians, while this statistic decreases to 10–20 % in high-income countries/regions such as the United States or European

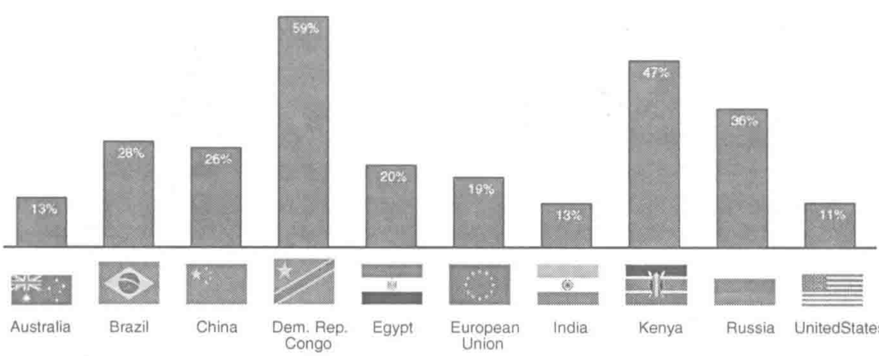

Fig. 1.3 Percentage of pedestrian deaths over all the traffic deaths in several regions of the world

Union. In the future, as the number of vehicles rises in low-income countries, the percentage of vehicle passengers will increase (as a result of having more drivers and occupants) with respect to the pedestrians.

In view of these terrible statistics, during the last twenty years automative companies and suppliers have progressively turned their safety efforts also to pedestrian protection. At the beginning, research was focused on optimizing the vehicle's physical parts in order to minimize impact severity. This research direction is often referred to as improving safety through design. Some examples are collapsing fenders, hood and windshield, or increasing the space between (a softer) hood and the engine to accommodate the pedestrian's head in the case of a crash. Later, more intelligent approaches were introduced. The first investigations on intelligent systems addressing pedestrian protection were conducted in the 1990s by Papageorgiou (MIT), Gavrila (Daimler), Broggi and Bertozzi (University of Parma). Nowadays, pedestrian safety has become an interesting research and development topic for companies, governments and research centers. Some examples of such interest can be seen in the last three European Union Programmes for research, also known as Framework Programmes (FP):

- Under 5th FP: PROTECTOR (4.4 million €, 2002–2003) and SAVE-U [268] (8 million €, until 2005), with Faurecia as coordinator, and CEA, Volkswagen AG, Daimler Chrysler AG, Siemens VDO Automotive AG and Mira Ltd as partners.
- Under 6th FP: PReVENT's APALACI [256] (3.75 million €, 2004–2007), coordinated by FIAT with partnering from Daimler Chrysler AG, Robert Bosch GmbH, Ibeo Automobile Sensor GmbH, Volvo and University of Parma.
- Under 7th FP: ADOSE (10.2 million €, 2008–2011), coordinated by FIAT; and FNIR [96] (3.12 million €, 2008–2010) coordinated by Autoliv AB.

Pedestrian Protection Systems (PPSs) are a particular type of ADAS focused to pedestrian safety. A PPS is formally defined as a system that detects both static and moving people in the vehicle's surroundings (typically in the front area) in order to provide information to the driver and perform evasive or braking actions on the host vehicle if needed. Pedestrian detection before the impact (either long or short term) is crucial given that the severity of injuries for the pedestrian decreases with the crashing vehicle's speed [11]. Hence, any reduction in the speed can drastically reduce the crash severity. According to [11], pedestrians have a 90 % chance of surviving to car crashes at 30 km/h or below, but less than 50 % chance of surviving to impacts at 45 km/h or above. The benefits of PPSs are twofold: (1) PPSs are able to minimize the reaction time with respect to the human drivers (in humans it depends on age, hours of continuous driving, alcohol consumption, distractors, day/night, etc.) and (2) they can control active measures such as airbags or brakes to minimize the potential impact.

1.4 The Role of Computer Vision

The central problem of PPSs corresponds to the task of detecting pedestrians. In order to detect objects (e.g., vehicles, pedestrians, obstacles) in the distance, ADAS use sensors that provide data to a computer/controller that processes them and performs the corresponding actions. The most widely used sensors for pedestrian detection are cameras mainly thanks to the rich high resolution information they provide, with cues such as edges, contours, texture or even relative temperature. In order to capture these cues, cameras working either in the visible or infrared spectrum are used.

A pedestrian detector must tackle several challenges, which are independent from the used sensor:

- A high variability in pedestrians' appearance since they can be of different size (especially in height), change their pose, wear different clothes and carry different objects. Moreover, pedestrians appear at different viewing angles (e.g., lateral and front/rear positions) and are imaged from a large range of distances (at least 25 m and up to 60 m, which corresponds to 60–15 pixel pedestrians, depending on the image acquisition system.
- Deal with cluttered background under a wide range of illumination and weather conditions, and in the presence of shadows and poor contrast. Furthermore, pedestrians can be partially occluded by vehicles, street furniture or other pedestrians.
- Manage highly dynamic scenes where not only pedestrians and other objects are in motion but also the camera is moving.
- Different temperatures and distances affect nighttime detection with infrared cameras.
- A high demanding performance in terms of accuracy (false alarms and misdetections) and system reaction time.

As can be appreciated, the topic differs from general human detection systems such as surveillance or human-machine interfaces. Although PPSs can indeed make use of the techniques developed for these applications, many typical simplifications as the following must be discarded attending to the inherent challenges of PPSs:

- Static camera assumption, common in surveillance, is not applicable. Hence, traditional techniques related to background subtraction cannot be directly applied in this research.
- Indoor illumination, common in human-machine interfaces, does not suit driving assistance applications.
- Model size is typically more constrained in dataset retrieval than in PPSs. For example, the human model in visual dataset searching is normally focused on well-seen people with a considerable amount of pixels to analyze. In the case of PPSs, pedestrians at 50 m can measure up to 10 pixels high. In the case of retrieval, however, the pose variability is more flexible than in PPSs, in which pedestrians are assumed to stand up on the road or pavement except in very rare cases or in the case of children [14].

The reader can refer to the surveys in [97] and [215], focused on generic human detection, and [187], focused on intelligent transportation systems, for more details about human detection for applications different from PPSs.

1.5 Generic Framework

A basic pedestrian detector is often based on two components: one that selects image regions likely to contain a pedestrian and another that classifies the regions as pedestrian or non-pedestrian. These two parts are usually called candidates generation and classification. Many object detection algorithms such as face or vehicle detection use this two-step approach. As it will be seen in further chapters, the development of new techniques not only has led to the improvements of these two components but also to the inclusion of additional ones. For example, since a PPS is focused on the continuous detection of pedestrians, it seems natural to include a new component capable of *following* the pedestrians through time. This would provide information on the direction of the targets, for instance. From here on, we will refer to these components as modules. Breaking down the systems into several modules is a convenient way to favor its understanding. For example, in Sun's vehicle detection review [286], techniques are divided into 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.

It is not easy to divide the PPSs proposed in the literature in specific modules: sometimes there is not a clear description of them, sometimes different modules are mixed together in a single algorithm, and even the functionality of a module is completely omitted. In order to provide an organized and comprehensive review of the existing techniques, we first describe a generic architecture of six different modules. Each module has its own objectives and responsibilities in the system, so by fitting each algorithm in the literature in its corresponding module it will be easier to compare and analyze the proposals. The used modules are the following:

- Preprocessing: it takes input data and prepares it to further processing. The data comes mainly from a camera but in some cases information is also acquired from car sensors, odometers, etc. The tasks carried out are diverse, some examples are sensors synchronization, adjust camera exposure time and gain, and calibration.
- Candidates Generation: it extracts regions of interest (candidates) from the image to be sent to the classification module avoiding as many *non-pedestrian* regions as possible.
- Classification: it receives a list of candidates to be classified as pedestrian or non-pedestrian.
- Verification and refinement: it verifies and refines the candidates classified as pedestrians, referred to as detections. The verification filters false positives using criteria not overlapped with the classifier while the refinement performs a fine

segmentation of the pedestrian (not necessarily silhouette-oriented) so to provide an accurate distance estimation or to support the following module, tracking.
- Tracking: it follows the detected pedestrians along time with several purposes such as avoiding spurious false detections, predict the next pedestrian position and direction and even other high-level tasks such as inferring pedestrian behavior.
- Application: it takes high level decisions (braking, steering, etc.) by making use of the information provided by the previous modules. This module represents a complete area of research, which includes not only driver monitoring but also psychological issues, human-machine-interaction, etc.

Figure 1.4 shows a schematic view of the PPS architecture. It is worth noting that although no feedback arrows have been included, several works make use of feedback or iterative cooperation between modules, e.g., classification and candidates generation.

An important aspect of the book is that it is focused on works using passive sensors, cameras working either in the visible (typically for daytime) or infrared (for nighttime) spectra, which 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 to the infrared either as NIR (near infrared, 0.75–1.4 μm) or FIR (far infrared, 6–15 μm).[1] 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 FIR sensors, they capture relative temperature, which is very convenient for distinguishing targets such as pedestrians or vehicles from asphalt or trees. For an analysis of radar, laserscanner or ultrasonic sensors please refer to [50].

1.6 Book Outline

This book surveys the state of the art in pedestrian detection for PPSs putting the emphasis on two aspects: (1) to overview the techniques used in PPSs from the first of these systems to the latest ones; (2) to provide a global viewpoint not specifically focused on classification techniques but also on the different ingredients of a complete PPS. The book can be seen as an extended version of the survey in [121] with updated references and comments. Note that in this survey we have deliberately avoided a quantitative performance assessment between the existing techniques, as this is out of the scope of the book. For surveys specifically evaluating this aspect for some selected proposals the reader can refer to the excellent surveys by Hussein et al. [152], Dollár et al. [74] and Enzweiler et al. [83]. Additionally, for a more generic survey on the pedestrian safety measures from a transportation viewpoint (not only on-board PPS but also infrastructures) the authors propose [109] as a the relevant publication to read.

[1] While it would be more precise to refer to this range as long wave or thermal infrared, in this book we use the term FIR (which in fact corresponds to the 15–1,000 μm range) given that this is the most common naming in the pedestrian detection literature.

此为试读，需要完整PDF请访问：www.ertongbook.com

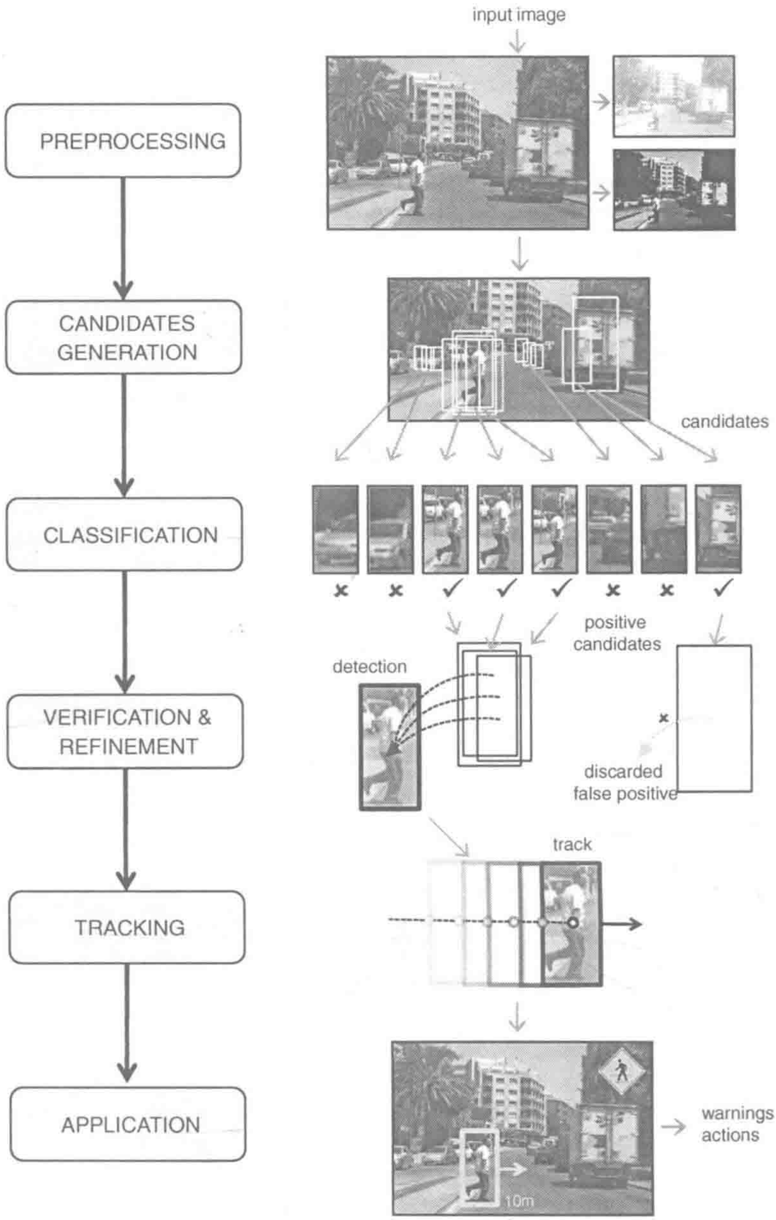

Fig. 1.4 Generic pedestrian protection system architecture with example results of each module. Note that in some proposals several modules are combined in a single one or can loop back the results between them, we have not included these case for the sake of clarity

The following two chapters are focused on the two core modules of the system: candidates generation in Chap. 2 and classification in Chap. 3, this latter being the longest of the book as it concentrates most of the literature. The rest of the system components (including preprocessing, verification and refinement, tracking and application), together with an overview of the aspects related to real-time, are surveyed in Chap. 4. Chapter 5 analyzes the existing methodologies for evaluating both classification and complete systems, including protocols and databases. Finally, in Chap. 6 we present the overall conclusions and the general perspective for the future of this research area.

Chapter 2
Candidates Generation

Candidates Generation extracts regions of interest (ROIs), usually rectangular windows, from the input image that are likely to contain a pedestrian avoiding as many non-pedestrian regions as possible. These ROIs will be sent to the classification module. Candidates generation is also referred to as foreground segmentation [125] or candidates extraction [4], though the most common name is candidates generation. It is the first core module of a PPS, and the approaches can be divided into 2D-based, 3D-based and motion-based.

2.1 2D-Based Approaches

The simplest and most extended candidates generation approach is the sliding windows algorithm. It is used in many object detection systems, especially in the ones in which there is barely any constraint on the candidates' position and size. It consists in exhaustively scanning the input image with window candidates only constrained by aspect ratio and perhaps position (e.g., avoiding the image's upper-part) but not by any complex reasoning (Fig. 2.1).

One of the first approaches using sliding windows is presented by Papageorgiou et al. [241]. The authors propose to scan the input image with windows of 128×64 pixels to be sent to the classifier. In order to achieve multi-scale detection, the authors propose to scan the image from 0.2 to 2.0 times its original size with increments of 0.1. This candidates generation procedure was popularized after the comprehensive study by Dalal in his PhD Thesis [59], but it is successfully used in many human detection systems, e.g., surveillance, indexing, sports video analysis, etc. There exist two different approaches to multi-scale sliding windows depending on what element is scaled: the image or the detection window. The former approach, which is used in [59, 83, 93, 189, 241, 299, 328], consists in an image-pyramid of s scales (typically from 8 [60] up to 50 [72]) that is scanned with a constant-size detection window. In the second approach the detection window is resized in s different

D. Gerónimo and A. M. López, *Vision-based Pedestrian Protection Systems for Intelligent Vehicles*, SpringerBriefs in Computer Science,
DOI: 10.1007/978-1-4614-7987-1_2, © The Author(s) 2014

Fig. 2.1 Sliding windows concept. A windows scans the input image in all the possible positions. In real systems, the resulting windows overlap and are of different sizes

scales to scan a fixed size input image [123, 245, 310]. Notice that the type of scanning limits the use of one feature or another, i.e., the local image features in a resizing detection window must be also resizeable, while in the image-pyramid scanning they are not restricted. These aspects, together with a description of intermediate scanning methods aimed at accelerating the detection process, will be detailed in Chap. 3.

Although sliding windows is successfully used in many different computer vision applications, in the case of PPSs this technique has two main drawbacks. First, the number of candidates is large, which makes it difficult to fulfill the real-time requirements. As an example, a regular scan over a 640×480 pixels image can provide from hundreds of thousands to millions of windows, depending on the sampling step and the minimum window size. An intuitive procedure is to decrease the sampling step to reduce the number of windows, however this also decreases the chance that a candidate window correctly frames the pedestrian in an unbiased manner for its further classification. Second, many irrelevant regions such as sky regions or windows inconsistent with perspective are likely to be sent to the classification, which increases the potential number of the system's false positives. An intuitive approach to reduce this kind of candidates is to discard the top 1/3 of the image. However, this solution does not avoid perspective-incorrect candidates such as small-sized windows in near road positions. A more sophisticated approach that depends on the classifier used but it is applicable in any context is proposed by Lampert et al. [174]. This proposal makes use of a branch-and-bound technique that bounds the classifier output, providing a globally optimal solution at sublinear time. The authors successfully implement the proposal using different classifiers, e.g., SVM using spatial pyramid kernels or χ^2-distance. However, as it will be seen later, there are more clever approaches that take advantage of the application's information.

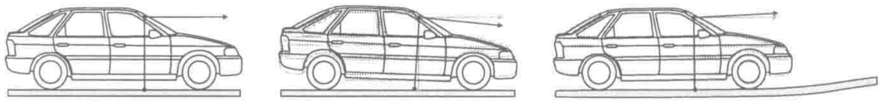

Fig. 2.2 Different situations in which the relative camera–road position varies

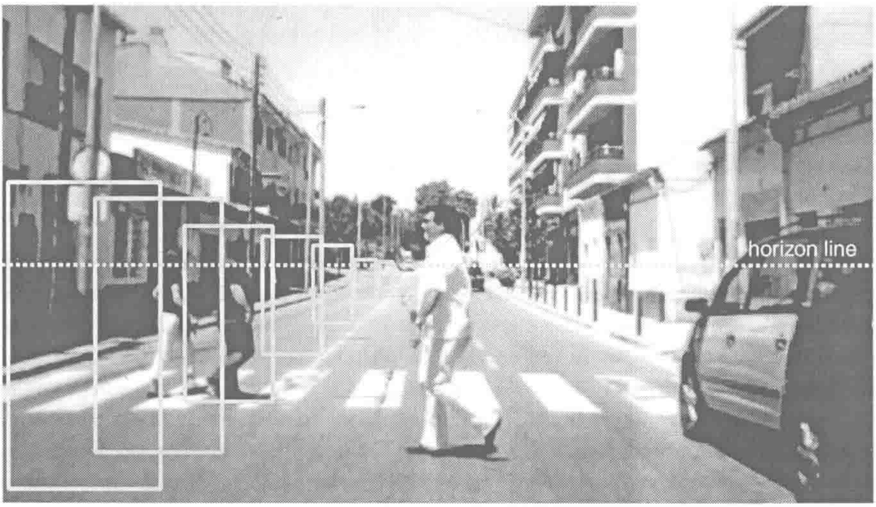

Fig. 2.3 Flat world assumption concept. The *horizon line* computed as the road position at infinity is one of the ways to represent the camera–road position

Pedestrian detection systems soon started to include some scene's prior knowledge aimed at improving sliding windows. The approach called *flat world assumption* defines the techniques that restrict the candidate search to the ground plane without any other information than the initial camera position with respect to the plane. It consists in fixing a constant camera-road position, represented as a fixed 3D plane or a horizon-line position, and then scanning the road with pedestrian-sized windows (see *Aligned Road-Horizon* in Fig. 2.2). Accordingly, the candidate windows will be projected in specific places of the image depending on their position and distance (see Fig. 2.3). By sampling just the estimated road the number of candidates is reduced to tens of thousands of windows, and again this number depends on the scan's density and the maximum detection distance. Different research groups, both in pedestrian [23, 113] and vehicle [254] detection made use of this idea. This technique performs significantly well when dealing with non-urban roads such as highways, where in general the road slope does not change much and in which the vehicle dynamics (e.g., braking, accelerating) are smooth. In order to solve the problem of changing camera-road position, Ponsa et al. [254] fix three different scanning road planes to detect vehicles in highways. Nevertheless, this is insufficient for pedestrian detection in urban scenes, since not only the target is smaller but also the road slope changes

此为试读，需要完整PDF请访问：www.ertongbook.com