# Introduction

Raja Bala<sup>1</sup> and Robert P. Loce<sup>2</sup>

<sup>1</sup> Samsung Research America, Richardson, TX, USA <sup>2</sup> Conduent Labs, Webster, NY, USA

With rapid advances in driver assistance features leading ultimately to autonomous vehicle technology, the automobile of the future is increasingly relying on advances in computer vision for greater safety and convenience. At the same time, providers of transportation infrastructure and services are expanding their reliance on computer vision to improve safety and efficiency in transportation and addressing a range of problems, including traffic monitoring and control, incident detection and management, road use charging, and road condition monitoring. Computer vision is thus helping to simultaneously solve critical problems at both ends of the transportation spectrum—at the consumer level and at the level of the infrastructure provider. The book aims to provide a comprehensive survey of methods and systems that use both infrastructural and in-vehicle computer vision technology to address key transportation applications in the following three broad problem domains: (i) law enforcement and security, (ii) efficiency, and (iii) driver safety and comfort. Table 1.1 lists the topics addressed in the text under each of these three domains.

1

This chapter introduces and motivates applications in the three problem domains and establishes a common computer vision framework for addressing problems in these domains.

# 1.1 Law Enforcement and Security

Law enforcement and security are critical elements to maintaining the well-being of individuals and the protection of property. Societies rely on law enforcement agencies to provide these elements. Imaging systems and computer vision are means to sense and interpret situations in a manner that can amplify the effectiveness of officers within these agencies. There are several common elements shared by computer vision law enforcement and security applications, such as the detection and identification of events of interest. On the other hand, there are also distinctions that separate a security application from law enforcement. For instance, prediction and prevention are important for security applications, while accuracy and evidence are essential for law enforcement. In many cases, modules and components of a security system serve as a front end of a law enforcement system. For example, to enforce certain traffic violations, it is necessary to detect and identify the occurrence of that event.

## 1

Computer Vision and Imaging in Intelligent Transportation Systems, First Edition.

Edited by Robert P. Loce, Raja Bala and Mohan Trivedi.

<sup>© 2017</sup> John Wiley & Sons Ltd. Published 2017 by John Wiley & Sons Ltd.

Companion website: www.wiley.com/go/loce/ComputerVisionandImaginginITS

Problem domains	Applications and methods	Imaging system employed
Law enforcement and security	License plate recognition for violations	Infrastructure
	Vehicle classification	Infrastructure
	Passenger compartment violation detection	Infrastructure, in-vehicle
	Moving violation detection	Infrastructure
	Intersection monitoring	Infrastructure
	Video anomaly detection	Infrastructure
Efficiency	Traffic flow analysis	Infrastructure
	Parking management	Infrastructure, in-vehicle
	License plate recognition for tolling	Infrastructure
	Passenger compartment occupancy detection	Infrastructure, in-vehicle
Driver safety and comfort	Lane departure warning	In-vehicle
	Collision avoidance	In-vehicle
	Pedestrian detection	In-vehicle
	Driver monitoring	In-vehicle
	Traffic sign recognition	In-vehicle
	Road condition monitoring	In-vehicle, infrastructure

Table 1.1 Taxonomy of problem domains and applications described in the book.

Consider the impact of moving vehicle violations and examples of benefits enabled by computer vision law enforcement systems. There is a strong relationship between excessive speed and traffic accidents. In the United States in 2012, speeding was a contributing factor in 30% of all fatal crashes (10,219 lives) [1]. The economic cost of speeding-related crashes was estimated to be \$52 billion in 2010 [2]. In an extensive review of international studies, automated speed enforcement was estimated to reduce injury-related crashes by 20–25% [3]. The most commonly monitored moving violations include speeding, running red lights or stop signs, wrong-way driving, and illegal turns. Most traffic law enforcement applications in roadway computer vision systems involve analyzing well-defined trajectories and speeds within those trajectories, which leads to clearly defined rules and detections. In some cases, the detections are binary, such as in red light enforcement (stopped or passed through). Other applications require increased accuracy and precision, such as detecting speed violations and applying a fine according to the estimated vehicle speed. There are other deployed applications where the violation involves less definitive criterion, such as reckless driving.

Several moving violations require observation into the passenger compartment of a vehicle. Failure to wear a seat belt and operating a handheld cell phone while driving are two common safety-related passenger compartment violations. Seat belt use in motor vehicles is the single most effective traffic safety device for preventing death and injury to persons involved in motor vehicle accidents. Cell phone usage alone accounts for roughly 18% of car accidents caused by distracted drivers [4]. In addition, the National Highway Traffic Safety Administration (NHTSA) describes other behaviors resulting in distracted driving, including occupants in the vehicle eating, drinking, smoking, adjusting radio, adjusting environmental controls, and reaching for an object in the car. The conventional approach to enforcement of passenger compartment violations has been through traffic stops by law enforcement officers. This approach faces many challenges such as safety, traffic

disruption, significant personnel cost, and the difficulty of determining cell phone usage or seat belt usage at high speed. Imaging technology and computer vision can provide automated or semiautomated enforcement of these violations.

Security of individuals and property is another factor in the monitoring of transportation networks. Video cameras have been widely used for this purpose due to their low cost, ease of installation and maintenance, and ability to provide rich and direct visual information to operators. The use of video cameras enables centralized operations, making it possible for an operator to "coexist" at multiple locations. It is also possible to go back in time and review events of interest. Many additional benefits can be gained by applying computer vision technologies within a camera network. Consider that, traditionally, the output of security cameras has either been viewed and analyzed in real-time by human operators, or archived for later use if certain events have occurred. The former is error prone and costly, while the latter has lost some critical capabilities such as prediction and prevention. In a medium-sized city with several thousand roadway cameras, computer vision and video analytics allow a community to fully reap the benefits of analyzing this massive amount of information and highlighting critical events in real-time or in later forensic analysis.

In certain security and public safety applications, very rapid analysis of large video databases can aid a critical life or death situation. An Amber Alert or a Child Abduction Emergency is an emergency alert system to promptly inform the public when a child has been abducted. It has been successfully implemented in several countries throughout the world. When sufficient information is available about the incident (e.g., description of captor's vehicle, plate number, and color), a search can be conducted across large databases of video that have been acquired from highway, local road, traffic light, and stop sign monitoring, to track and find the child. Similar to Amber Alert and much more common is Silver Alert, which is a notification issued by local authorities when a senior citizen or mentally impaired person is missing. Statistics indicates that it is highly desirable that an Amber-/ Silver Alert-related search is conducted in a very fast and efficient manner, as 75% of the abducted are murdered within the first 3h. Consider a statement from the US West Virginia code on Amber Alert 15-3A-7:

The use of traffic video recording and monitoring devices for the purpose of surveillance of a suspect vehicle adds yet another set of eyes to assist law enforcement and aid in the safe recovery of the child.

Human analysis of video from thousands of camera could take many days, while computer vision methods have the potential to rapidly extract critical information. The speed can be scaled by the available computational power, which is rapidly advancing due to high-speed servers and cloud computing.

Whether it is safety of individuals or security of property, recognition of a vehicle is a key component of a roadway security system. Vehicles traveling on the public roadways in most countries are required by law to carry a clearly visible placard with a unique identifier that is registered with the local government. This placard (license plate) can contain various symbols-letters, numbers, logos, etc.-based on local government regulations and the vehicle class. Given the common requirement for its presence and ease of visibility, the license plate has become the default means for identifying a vehicle and/or its registered operator. Automated license plate recognition (ALPR), also referred to as automated number plate recognition (ANPR), leverages computer vision algorithms to extract license plate information from videos or still images of vehicles. ALPR has become a core technology within modern intelligent transportation systems. Surveillance and police enforcement applications leverage ALPR systems to provide real-time data gathering to support law enforcement efforts. For

example, a 2012 study [5] conducted on the usage of ALPR technology by police agencies found that 80% of larger agencies (those with 1000+ sworn officers) were leveraging ALPR in some way. Results from this same study also indicated that police agencies using ALPR reported that the use of technology had increased the recovery of stolen vehicles by 68% and overall arrests by 55%.

A trend that favors improved law enforcement and security in transportation settings is the increasing intelligence across broad camera networks. Cities such as London and Tokyo are said to have over 500,000 government cameras, while Chicago has linked private, school, roadway, and police cameras into a massive interconnected network. Several factors are driving this trend. There is the fight against large-scale terrorism and crime with limited human resources. Camera and network technology are continually becoming more capable and less costly. Cloud computing is also enabling a flexible, scalable architecture for big data analysis. Computer vision becomes the intelligent connecting element that enables the camera network and computing resources to address the societal need.

## 1.2 Efficiency

The efficiency of a roadway network impacts expended time of individuals, fuel usage, and pollution all of which are key factors contributing to the quality of life of a community. Roadway imaging with computer vision is being increasingly applied to optimize efficiency. Example applications include traffic flow analysis, video-based parking management, open road tolling, and high-occupancy lane management.

Traffic is the movement of people and goods in a public space, where the movement may involve a car, public transport vehicle, bicycle, or foot travel. Data derived from traffic volume studies can help local governments estimate road usage, volume trends, and critical flow time periods. These estimates can be used to optimize maintenance schedules, minimize conflicts, and optimize the timing of traffic enforcement. Real-time traffic flow data can also enable efficient incident management, which consists of incident detection, verification, and response. Traffic variables of interest are flow, speed, and concentration with respect to road capacity. Each of these variables involves detection and recognition and tracking of a particular traveling entity (pedestrian, bicycle, car, etc.) Manual methods involve monitoring a region onsite or via a video feed, both of which are very labor intensive and tend to provide a limited snapshot of information. The past decade has seen a trend toward increased automation, leveraging the ubiquity of roadway cameras and advances in computer vision toward vehicle detection, vehicle classification, and pedestrian detection. This trend is bringing fine-grain and persistent analysis to traffic engineers and, in turn, increasing the efficiency of our roadways.

Urban parking management is receiving significant attention due to its potential to reduce traffic congestion, fuel consumption, and emissions [6, 7]. Real-time parking occupancy detection is a critical component of parking management systems, where occupancy information is relayed to drivers in real time via smartphone apps, radio, the Internet, on-road signs, or GPS auxiliary signals. This can significantly reduce traffic congestion created by vehicles searching for an available parking space, thus reducing fuel consumption. Sensors are required to gather such real-time data from parking venues. Video-based sensing for monitoring on-street parking areas offers several advantages over sensors such as inductive loops, ultrasonic sensors, and magnetic in-ground sensors. One advantage is that one video camera can typically monitor and track several parking spots (see Figure 1.1), whereas multiple magnetic sensors may be needed to reliably monitor a single parking space. Another advantage is that device installation and maintenance is less disruptive in the case of video cameras when compared to in-ground sensors. Video cameras can also support other tasks such as traffic law enforcement and surveillance since they capture a wider range of useful visual



Figure 1.1 Examples of a camera view sensing multiple parking stalls.

information including vehicle color, license plate, vehicle type, speed, etc. A video-based parking occupancy detection system can, therefore, provide a convenient, cost-effective solution to the sensing task and also provide additional functionality for traffic law enforcement and surveillance.

A third problem is highway congestion. Government officials and members of the transportation industry are seeking new strategies for addressing the problems associated with high traffic volumes. One such mechanism to reduce the congestion on busy highway corridors is the introduction of managed lanes such as high-occupancy vehicle (HOV) lanes that require a minimum number of vehicle occupants and high-occupancy tolling (HOT) lanes that set a tolling price depending upon the number of occupants. Due to imposed limitations and fees, HOV/HOT lanes are often much less congested than other commuter lanes. However, the rules of the HOV/HOT lane need to be enforced to realize the congestion reducing benefits. Typical violation rates can exceed 50-80%. Current enforcement practices dispatch law enforcement officers to the roadside to visually examine passing vehicles. Manual enforcement can be a tedious, labor-intensive practice, and, ultimately, ineffective with enforcement rates of typically less than 10% [8]. Besides enduring environmental conditions of snow, darkness, sunlight reflections, and rain, law enforcement officers also have to deal with vehicles traveling at high speeds that may have darkened/tinted glass, reclining passengers, and/or child seats with or without children. As a result, there is a desire to have an automated method to augment or replace the manual process. Practical imaging-based systems have been demonstrated using nearinfrared (NIR) illumination and multiple cameras triggered by induction loops or laser break-beam devices (Figure 1.2).

## 1.3 Driver Safety and Comfort

Many vehicles today employ a so-called *Advanced Driver Assistance System* (ADAS) that uses cameras and various sensors to make inferences about both the driver and the environment surrounding a vehicle. An important element of ADAS is lane departure warning. More than 40% of all fatal roadway accidents in 2001 involved a lane or road departure [9], resulting primarily from driver distraction, inattention, or drowsiness [10]. *Lane Departure Warning* (LDW) systems [11] track roadway markings using a video camera mounted near the rearview mirror or on the dash board of a vehicle so the area in front of the vehicle may be viewed. A warning signal is given to the driver if a vehicle unintentionally approaches a lane marking (i.e., without activating a turn signal). The prevalence of LDW systems is expected to rapidly increase, with various tax incentives being proposed in the United States and Europe for vehicles with LDW systems. LDW algorithms face the daunting task of operating in real-time and under multifarious weather conditions, in order to detect and decipher within this limited field of view a wide assortment of lane markings.



Pedestrian detection is another important element of ADAS. While this problem has been extensively studied from the viewpoint of fixed video surveillance cameras, many new challenges arise when a camera is mounted on a common moving vehicle. The detection must comprehend a wide range of lighting conditions, a continuously varying background, changes in pose, occlusion, and variation in scale due to the changing distance. Technologies available today for pedestrian detection use some combination of appearance- and motion-based techniques. A generalization of this problem is collision avoidance with other vehicles, bicyclists, and animals. Feedback can include direct control on vehicle motion, with popular examples being adaptive cruise control and automatic braking.

A third form of ADAS directly monitors the driver's attention with the use of a driver-facing camera and possibly other sensors placed in contact with the driver's body. A significant challenge is making accurate inferences on the state of the driver by monitoring video of facial expression, eye movement and gaze in the presence of vehicle motion, varying illumination, and large variation in affective expression across humans. A promising direction in this domain is the use of adaptive and online learning techniques whose inferences are personalized to a specific driver and vehicle.

A fourth class of ADAS monitors relevant aspects of the static environment, such as road conditions and traffic signs. The European ASSET program has been actively pursuing within-vehicle camera-based methods to detect the friction (conversely, slipperiness) of roads, a significant factor in fatalities worldwide. Imaging systems using polarization filters and IR imaging are being explored to address this problem. Traffic signs provide crucial information about road conditions, and can often be missed by the driver due to weather, occlusion, damage, and inattention. A computer vision system that can quickly and accurately interpret traffic signs can be an invaluable aid to the driver. Challenges pertain to poor image quality due to inclement weather, vehicle motion, the tremendous variety in the scene and objects being captured, and the need for real-time inferences. ADAS can be classified by the degree of autonomy that is enabled during active driving. In the United States, the NHTSA has defined vehicle automation as having five levels [12] as follows:

- *No-Automation* (Level 0): The driver is in complete and sole control of the primary vehicle controls—brake, steering, throttle, and motive power—at all times.
- *Function-Specific Automation* (Level 1): Automation at this level involves one or more specific control functions. Examples include electronic stability control or precharged brakes, where the vehicle automatically assists with braking to enable the driver to regain control of the vehicle or stop faster than possible by acting alone.
- *Combined Function Automation* (Level 2): This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering.
- *Limited Self-Driving Automation* (Level 3): Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time.
- *Full Self-Driving Automation* (Level 4): The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles.

Full Self-Driving Automation is embodied in what is often referred to as autonomous vehicles, a much publicized example being the Google driverless car. As of September 2015, Google's fleet of autonomous vehicles have logged 1,200,000 driverless road miles. A key technology used in Google's implementation is a roof-mounted Velodyne 64-beam laser, which creates a three-dimensional (3D) map of the surrounding environment in the immediate area of about 50 ft. The 3D image is combined with high-resolution maps that have been programmed into the vehicle's control system. The laser system can differentiate among a large variety of objects, including cars, pedestrians, cyclists, and small and large stationary objects. Four radars (one for front, back, left, and right) sense any fastmoving objects at distances further than the laser detection range and are used to give the car farsighted vision for handling high speeds on freeways. A front-mounted camera and computer vision algorithms analyze the road ahead of the car, observing road signs and stop lights for information that a human driver typically uses. Other sensors include a GPS, an inertial measurement unit, and wheel encoder. While autonomous vehicle technology poses many challenges to current roadway legislation, it does offer great potential to mobilize citizens with impairments and could make driving safer due to comprehensive sensing and rapid decision making. For further details, the reader is referred to the recent IEEE Spectrum Online article [13].

## 1.4 A Computer Vision Framework for Transportation Applications

Computer vision can be broadly described as the task of interpreting and making sense of the world around us from images and video. There are many levels and types of interpretation, beginning at the basic level with detecting and recognizing individual objects and their motion, and evolving onto higher levels of inference on interactions and relationships between multiple objects, human–object interactions, and semantic reasoning of human behavior. As such, computer vision falls within the



Figure 1.3 Computer vision pipeline for transportation applications.

broader domain of artificial intelligence (AI), and can be thought of as the branch of visual sensing, perception, and reasoning within the field of AI. In the context of transportation systems, computer vision enables the interpretation of the environment outside of and within automotive vehicles, whether it is for the purpose of enhancing safety, efficiency, or law enforcement. In many instances, the computer vision system augments, or if reliable enough, even replaces human interpretation so as to reduce errors, cost, effort, and time. A basic computer vision pipeline is shown in Figure 1.3. Each of the blocks is described next.

### 1.4.1 Image and Video Capture

The pipeline begins with an imaging system comprising one or more cameras that capture digital images or video of a relevant scene. Infrastructural cameras may be installed on bridges, gantries, poles, traffic lights, etc. In-vehicle cameras are positioned either outside the vehicle, providing front and rearviews of the environment surrounding the vehicle, or within the vehicle usually for the purpose of monitoring the driver's alertness and attention. Factors that are considered in designing the imaging system are field of view, coverage of spectral bands, spatial and temporal resolution, form factor, ruggedness, and cost. A stereo- or multicamera network often requires establishing the relationship among the camera views, and between camera coordinates and 3D world coordinates. Applications utilizing stereo vision techniques in the book include moving violation detection, pedestrian detection, and parking management. Also to be considered are the bandwidth and cost of the data transmission technology used to send video data from the camera to the subsequent computational engines which may be logically and physically co-located with or separated from the imaging system. Another practical factor is the electrical power source. In urban areas, power is typically supplied via lines that support traffic controls and lighting, while more remote regions sometimes harvest energy locally using photovoltaics. The reader will be exposed to different types of imaging systems throughout the course of the book.

### 1.4.2 Data Preprocessing

The next step in the pipeline is data preprocessing. Well-known image processing operations such as brightness and contrast normalization, distortion compensation, noise filtering, spatial resizing, cropping, temporal frame rate conversion, and motion stabilization may be carried out to prepare

the images or video for subsequent analysis. The operations and their parameters are carefully tuned using knowledge of the characteristics of the imaging system, as well as the requirements of the computer vision task at hand. For example, images of vehicle license plates taken with infrastructure cameras require normalization for accurate performance in the presence of widely varying contrast due to ambient lighting and weather conditions. On the other hand, driver-facing video captured within a moving vehicle will likely have to undergo motion deblurring and stabilization due to both vehicle motion and the motion of the driver's head relative to the camera. Another common type of preprocessing identifies one or more regions of interest (ROIs) for subsequent interpretation. An example is human face detection, which is used for driver monitoring and in-vehicle passenger compartment violation detection. A second example is license plate localization, a precursor to plate recognition. Interestingly, such ROI detection can itself be viewed as a computer vision subproblem, involving its own feature extraction and decision operations. Edge detection is a very common preprocessing step for computer vision, and will be encountered in several applications, including lane detection and tracking (Chapter 11). Each chapter discusses data preprocessing unique to a specific problem domain. However, readers who are entirely unfamiliar with the field of image processing are encouraged to review the text by Gonzalez and Woods [14] for an excellent introductory overview of the subject.

Another type of preprocessing involves calibrating for camera characteristics. Two common forms are colorimetric and geometric calibration. Colorimetric calibration transforms the (typically RGB) camera color representation of the input image into a standardized color space such as sRGB, XYZ, or CIELAB. Parameters required to perform such a calibration include knowledge of the spectral sensitivities of the RGB filters, spectral distribution of the incident illumination, and optoelectronic transfer functions relating digital camera values to luminance at each pixel. Colorimetric calibration is important for tasks wherein color provides a critical cue in the inference. An example is locating features within the driver's face, wherein human skin color can serve as a cue. Geometric camera calibration involves relating camera pixel coordinates to real 3D world spatial coordinates. This problem is treated in detail in Chapter 5 describing moving violations, wherein relative motion within the coordinates of the video signal must be translated to absolute motion and speed in world coordinates.

#### 1.4.3 Feature Extraction

The third and fourth steps in the pipeline, namely feature extraction and decision inference, jointly define the computer vision component underpinning the overall system. Feature extraction is the process of identifying from the raw pixel data a set of quantitative descriptors that enable accurate interpretations to be made on the image. Essentially, feature extraction is a mapping from the original set of (possibly preprocessed) image pixels I to an alternate representation x in a suitably defined feature space. One benefit of feature extraction is dimensionality reduction; that is, the dimensionality of x is typically much smaller than that of I, and as such retains only the information that is germane to the inference task at hand. Numerous feature descriptors have been developed in the literature and used successfully in practical applications; the optimal choice is made based on knowledge of the domain, the problem, and its constraints. Feature descriptors fall into two broad categories: global and local features. Global features holistically describe the entire image or video, while local features represent a spatially or spatiotemporally localized portion of the image or video. A simple example of a global feature is a color histogram of an image. Such a feature may be employed for example in a vehicle identification or parking space occupancy detection problem, where color is an important cue. On the other hand, local features represent some spatiotemporally localized

attribute such as gradient, curvature, or texture in the image. A common example is the histogram of oriented gradients (HoGs) which will be encountered in a variety of transportation problems, including vehicle classification (Chapter 3), traffic sign recognition (Chapter 14), and pedestrian detection (Chapter 10). The scale-invariant feature transform (SIFT) and Haar descriptors are also commonly used for vehicle classification and other tasks. Note that global features can be modified to operate locally; for example, one can compute color histograms of  $N \times N$  image subblocks and concatenate them to form a composite feature *x* that captures a spatially varying color distribution within the image. Local descriptors may also be aggregated into a pooled feature describing the entire image, using techniques such as Bag-of-Words and Fisher Vector encoding, as will be encountered in one approach to passenger compartment violation detection in Chapter 4.

#### 1.4.4 Inference Engine

The inference stage takes as input the feature descriptors, and emits a hypothesis or decision. The inference operation can be thought of as a mapping y = h(x) that predicts an outcome y from input features x under a hypothesis  $h(\cdot)$ . Parameters of  $h(\cdot)$  are learned offline in a training phase using a set of data samples that are representative of what is to be encountered in the final application. As an example, if the inference task is to infer from an image of a vehicle whether a passenger compartment is occupied or vacant, then  $y \in \{y_0, y_1\}$  is a binary class label corresponding to either the "occupied" or "vacant" outcome, and  $h(\cdot)$  is a binary decision function whose parameters are learned to minimize the classification error on a set of training images of occupied and vacant compartments. If the task is to interpret a license plate, then features from individual character segments are processed through a character recognition algorithm that outputs one of 36 alphanumerical class labels,  $y \in \{y_1, ..., y_{36}\}$ . Hypothesis  $h(\cdot)$  is similarly learned from training samples of labeled license plate character images. If the task is to predict the speed of a vehicle, then y is a continuous valued variable, and may be predicted using, for example, a linear regression technique.

There are two broad categories of inference techniques using generative and discriminative models, respectively. Generative models attempt to describe the joint statistical distribution p(x, y) of input features and prediction outcomes. Since the joint distribution is often a complex, high-dimensional function, simplifying assumptions are made to arrive at computationally tractable solutions. Commonly, a Bayesian framework is employed, whereby the joint statistics are factored into a conditional likelihood function p(x|y), posterior distribution p(y|x), and a prior distribution p(y) for the output variables, each of which are modeled by tractable parametric forms. At the heart of a generative model is the search for the most likely outcome y given the observed features x. That is, the optimum y maximizes the posterior distribution:

$$y = h(\mathbf{x}) = \arg\max p(\mathbf{y} | \mathbf{x}) = \arg\max p(\mathbf{x} | \mathbf{y}) p(\mathbf{y})$$
(1.1)

The last expression in Equation 1.1 is arrived at with Bayes rule. An example of a generative model that is frequently encountered in the book is the Gaussian mixture model (GMM), used in traffic flow analysis, video anomaly detection, driver monitoring, parking space occupancy detection, and other applications. For an excellent introduction to Bayesian generative inference techniques, the reader is referred to Ref. [15].

Discriminative models, on the other hand, attempt to learn p(y|x) or the inference mapping h(x) directly from training data without explicitly characterizing the joint probability distribution. Examples of such models include linear regression, logistic regression, support vector machines (SVMs), and boosting techniques. Various discriminative approaches are encountered in the book in the chapters on parking management, traffic sign recognition, and driver monitoring.

Recently, deep learning methods have gained popularity particularly in various image classification tasks. The basic concept behind deep learning is that the features themselves are learned from the image data to optimally perform a specific classification tasks. Thus, rather than parsing the computer vision task in two phases of feature extraction and inference, deep methods perform the task in one stage, learning both the features and the patterns needed to make inferences. For this reason, these methods are also referred to as representation learning techniques, and have been shown to significantly outperform traditional techniques in certain large-scale image classification tasks. A practical advantage of deep learning over traditional techniques is not having to manually or heuristically design features for a given task and domain. Chapter 10 provides an excellent tutorial and application of deep learning to the pedestrian detection problem.

The chapters in the book introduce a broad variety of machine learning and inference models addressing the various problem domains. The reader will appreciate that the choice of inference technique depends on many factors, including the required accuracy, computational cost, storage and memory requirements, and the cost, labor, time, and effort required to train the system. Those seeking a more basic background on machine learning are referred to Ref. [15] for a fundamental treatment of the subject. Finally, while the focus of the text is on camera-based methods, practical systems often involve input from other non-visual sensors. One example is in-vehicle safety systems that utilize technologies such as radar, LiDAR, and motion sensors to monitor driver and vehicle behavior. Another example is parking management, discussed in Chapter 8, where magnetic, radio frequency identification (RFID), laser, and other sensors are commonly utilized to determine parking space occupancy. Where applicable, discussions are presented on how video-based systems compete with or are complementary to these nonvisual modes.

### 1.4.5 Data Presentation and Feedback

This is the final step in the computer vision pipeline. The inference engine outputs a prediction of an outcome in the form of a class label probability, or measure of some desired quantity. This outcome must then be acted upon and communicated to the relevant party, be it the driver, law enforcement agent, or urban planner. As would be expected, the method and format of presentation and feedback is strongly dependent on the task and application. In problems relating to driver safety, where the majority of computation takes place in the vehicle, inferences must be communicated in real time to the driver in order to allow immediate action and intervention. Built-in driver assistance systems employ various feedback mechanisms, including visual display on the dashboard, audio warnings, and tactile feedback on the steering wheel. An important consideration here is the design of user interfaces that provide timely feedback without distracting the driver, and that integrate seamlessly with other familiar functions such as entertainment and navigation. Systems that are more proactive can directly alter vehicle motion without driver involvement. Examples are adaptive cruise control, automatic braking in collision avoidance systems and ultimately the driverless vehicle.

Applications relating to transportation efficiency warrant a different type of data visualization. In a video tolling application, the output of a license plate recognition system may be sent to a tolling agency to initiate billing. Additionally, for those plates where automatic recognition has low confidence, the localized plate images may be sent for human interpretation. For an urban planning application, traffic flow metrics generated by the computer vision system may be presented on a dashboard to local agencies, who can then interpret and integrate the analysis into future road planning and optimization efforts. In a parking management system, the vehicle license plate and duration of parking are the relevant metrics that must be made available to the parking administrator for billing purposes. Parking availability at a given location is a metric that must be continuously measured and made available to parking authorities and drivers.

In the domain of law enforcement, image capture is performed within a transportation infrastructure such as a gantry or police vehicle. The computer vision pipeline is often executed on a centralized server, and various data feedback and presentation mechanisms are invoked. For example, a law enforcement officer may want to see a relevant metric such as vehicle speed or occupancy, along with a picture or video clip of the vehicle that can be used for adjudicating a traffic violation, as well as for evidentiary purposes.

One trend that must be recognized and leveraged in the step of data presentation is the ubiquity of mobile technologies such as smartphones, tablets, smartwatches, and other wearable devices that are intimately connected to all human agents in a transportation ecosystem. For example, today it is customary for a smartphone to be connected via Bluetooth to the audio system in the vehicle, or for drivers to use the GPS on their mobile device for driving directions. It is expected that the automobile of the future will seamlessly integrate multimedia information from mobile devices with visual interfaces on the dashboard to present a personalized and context-sensitive driving experience.

We conclude the chapter by observing that while each module in Figure 1.3 can be individually designed and tuned, it is highly beneficial to co-optimize the entire framework at a system level, so as to account for the interplay and interaction among the modules. For example, the choice of features and the design of the inference algorithm are closely coupled, and ideally must be co-optimized for a given computer vision task. Similarly, the choice of imaging system can strongly affect the optimum choice of features and algorithms, as seen in numerous examples in the text. In Chapter 13 on driver monitoring, the choice of data preprocessing and feature extraction method depend on whether active or passive illumination is used within the vehicle. In Chapter 4 on passenger compartment violation detection, cameras capturing single versus multiple spectral bands are compared in terms of detection accuracy. Similarly, Chapter 10 illustrates how thermal infrared cameras can simplify the pedestrian detection task. Also, as mentioned earlier, road condition monitoring benefits greatly from the use of thermal IR imaging and optical polarization. Finally, the means and requirements for information presentation and feedback must be comprehended while designing the preceding steps in the framework.

## References

- 1 Traffic Safety Facts, 2012 Data, Speeding. National Highway Traffic Safety Administration, NHTSA's National Center for Statistical Analysis, Washington, DC, DOT HS 812 021, 2014.
- 2 The Economic and Societal Impact Of Motor Vehicle Crashes, 2010 (Revised), NHTSA's National Highway Traffic Safety Administration, National Center for Statistical Analysis, Washington, DC, DOT HS 812 013, 2010.
- 3 L. Thomas, R. Srinivasan, L. Decina, L. Staplin, Safety Effects of Automated Speed Enforcement Programs: Critical Review of International Literature, *Transportation Research Record: Journal of the Transportation Research Board*, 2078, 117–126, 2008.
- 4 Distracted Driving 2009, Report, U.S. DOT National Highway Traffic Safety Administration, DOT HS 811 379, September, 2010. http://www-nrd.nhtsa.dot.gov/Pubs/811379.pdf (accessed September 9, 2016).
- **5** D. Roberts, M. Casanova, *Automated License Plate Recognition (ALPR) Systems: Policy and Operational Guidance for Law Enforcement,* Washington, DC: U.S. Department of Justice, National Institute of Justice, 2012.
- **6** D. C. Shoup, *The High Cost of Free Parking*, Chicago: Planners Press, American Planning Association, 2005, 7.

- 7 D. C. Shoup, Cruising for Parking, Transport Policy, 13(6), 479-486, 2006.
- **8** S. Schijns, P. Mathews, A Breakthrough in Automated Vehicle Occupancy Monitoring Systems for HOV/HOT Facilities, 12th International HOV Systems Conference, Houston, TX, April 20, 2005.
- **9** D. P. Jenkins, N. A. Stanton, G. H. Walker, M. S. Young, A New approach to Designing Lateral Collision Warning Systems, *International Journal of Vehicle Design*, 45(3), 379–396, 2007.
- 10 I-Car Advantage, Lane Departure Warning Systems, September 6, 2005.
- 11 C. Visvikis, T. L. Smith, M. Pitcher, R. Smith, Study on Lane Departure Warning and Lane Change Assistant Systems, Transport Research Laboratory Project Rpt PPR 374. http://www.trl.co.uk/ umbraco/custom/report\_files/PPR374.pdf and http://www.trl.co.uk/reports-publications/trl-reports/ report/?reportid=6789 (accessed October 13, 2016).
- 12 U.S. Department of Transportation Releases Policy on Automated Vehicle Development. National Highway Traffic Safety Administration. 30 May 2013. Retrieved December 18, 2013.
- 13 http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-carworks (accessed September 9, 2016).
- 14 R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 3rd Ed., Upper Saddle River, NJ: Prentice Hall, 2008.
- 15 C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2007.