VŠB – Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Computer Science

ALGORITHMS FOR IMAGE ANALYSIS IN TRAFFIC SURVEILLANCE SYSTEMS

TOMÁŠ FABIÁN

A dissertation submitted in partial fulfillment

of the requirements for the degree

Doctor of Philosophy in Computer Science

Field of study: 1801V001 - Informatics
Supervisor: doc. Dr. Ing. Eduard Sojka

Ostrava, 2014

To all who supported me all the way

Per aspera ad astra.

—Lucius Annaeus Seneca, Roman Philosopher, 4 BC – 65 AD

Condicio sine qua non.

—Anicius Manlius Torquatus Severinus Boëthius, Medieval Philosopher, 480 AD – 524 AD

PREFACE

The presence of various surveillance systems in many areas of the modern society is indisputable and the most perceptible are the video surveillance systems. These vision-based systems cover wide range of applications like safety, security preventing against dangerous situations, forensic tool, crime and accident precaution, object tracking, monitoring real-time industrial processes, and environmental monitoring. In order to automate all the aforementioned processes, sophisticated image analysis algorithms need to be carried out.

The history of surveillance cameras dates back to as early as 1965. Surveillance footage appeared with the introduction of inexpensive tapes and began to be used as evidence. The most important event in the development of this branch of industry was introduction of the charged coupled device (CCD) cameras enabling footage in low light and at night. With the continuously increasing computation power of specialized DSP circuits and general computers more and more tremendous human monitoring started to be superseded by autonomous intelligent surveillance systems.

This thesis mainly describes novel algorithm for vision-based estimation of the parking lot occupancy and the closely related topics of preprocessing of images captured under harsh conditions. The developed algorithms have their practical application in the parking guidance systems which are still more popular. One part of this work also try to contribute to the specific area of computer graphics denoted as direct volume rendering (DVR).

ACKNOWLEDGMENTS

CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

The aim of the dissertation thesis is to present the actual state of the algorithms and methods applied in the vision-based traffic surveillance systems and to propose the concept of the solution of the problems encountered in the specific area of the parking guidance systems. Primarily, we will focus on the algorithms for parking space occupancy detection and the closely related areas like motion detection and object tracking in the parking lot environment. The dissertation thesis contains, as the main part, a new discriminative classifier of parking lot occupancy devised on the theoretical foundations of random fields. The practical relevance of this research is encouraged by the interest of automotive industry.

Visual surveillance systems include a wide range of the computer vision related areas. Some of the most significant areas are motion detection, moving object classification, tracking, activity understanding and semantic description. Typical applications include traffic surveillance, security, classification of activities and behaviours, the pre-crash safety system of a vehicle, and various commercial applications like license plate recognition, toll road or speeding detection systems. In addition to these common usages, there are also more exotic applications like a Martian dust devil tracking. The Jet Propulsion Laboratory, which works closely with NASA, is investigating the abundance of the dust devils as they moves across the surface of planet Mars in Argyre Planitia and Hellas Basin [38]. Relevance of autonomous visual systems shows up also in the case of recent Curiosity rover mission. The mission uses radio relays via Mars orbiters as the principal means of communication between Curiosity and the Deep Space Network of antennas on Earth with a round-trip communications delay of nearly 30 minutes. The new on-board software enables autonomous navigation giving the rover autonomous hazard avoidance capabilities during the drives.

1.1 VISUAL SURVEILLANCE SYSTEMS

Video-based surveillance systems had evolved into sophisticated systems during the last 70 years. One of the very first closed circuit television (CCTV) systems was installed for observing the launch of the infamous ballistic missile V-2 at Peenemünde, Germany in 1942. Since that time, CCTV systems had successfully spread through many various application areas including monitoring dangerous industrial processes, security systems in banks, streets, stores, systems supporting transport safety and traffic surveillance. The first generation of the traditional analogue cameras connected by coaxial cables was replaced by the second generation in the early 1990s. Replacing the videotape recorder with a digital recorder (DVR) was the first step in the digital transition of video surveillance. These hybrid systems evolved into completely IP-based third generation devices and all of their components are digital and all transmissions are done by IP protocol [68]. There is also the constant discussion about the loss of privacy and liberty of people under surveillance. Prophetical Orwell's dystopia novel Nineteen Eighty-Four was first published in 1949 and introduced the Big Brother phenomenon. We could safely say that it has come true now. For example, the Great Britain is recognized as a leading user of CCTV and has roughly 4 million cameras installed in the streets, i.e. one for every 15 people living in the country. It is also interesting that only a small fraction of installed cameras are ever watched. According the survey in [49], the ratio of camera to screen is between 1:4 and 1:78. This implies that the records are often watched only in case of incident or emergency. Every technology can be potentially abused and visual surveillance is no exception. On the opposite side, the camera systems play a very important role in gathering evidence of criminal activity and can be effectively used to protect against vandalism.

Surveillance systems can be designed at various abstraction layers [106]. In some cases, we need only a simple motion detection, and sometimes sophisticated analysis involving classification and even explanation is required. The overall capabilities of designed system depend largely on sensors used for vehicle detection. The sensors can be divided into in-road and over-roadway sensors [108]. Intrusive in-road sensors are embedded in the pavement of the roadway or otherwise attached to the surface of the roadway. Examples of in-roadway sensors include inductive loop detectors, which require saw cuts in the pavement, weigh-in-motion sensors,

which are directly embedded in the pavement, magnetometers, which may be embedded or placed underneath a paved roadway or bridge structure, and tape switches, micro loops, pneumatic road tubes, and piezoelectric cables, which are mounted on the roadway surface. These sensors generally represent applications of mature technologies. The drawbacks to their use include disruption of traffic for installation and repair and failures associated with installations in poor road surfaces and use of substandard installation procedures. Non-intrusive over-roadway sensor is one that is mounted above the surface of the roadway either above the roadway itself or alongside the roadway, offset from the nearest traffic lane by some distance. Examples of over-roadway sensors are machine vision-based sensors that utilize cameras mounted on poles adjacent to the roadway, on structures that span the roadway, or on traffic signal mast arms over the roadway, microwave radar sensors mounted adjacent to the roadway or over the lanes to be monitored, ultrasonic, passive infrared, and laser radar sensors normally mounted over the lanes to be monitored (some passive infrared models can be mounted adjacent to the roadway), and passive acoustic sensors mounted adjacent to the roadway. Utilization of RFID for vehicle detection is described in [109]. It provides the means for low cost installation and maintenance as well as possible detection at high speed. The inherent drawback is the need to place transponders in every vehicle [77]. In this thesis, we are exclusively interested in vision-based surveillance technology bringing great flexibility, wide variety of output data and large area of coverage.

Many of the stated problems are commonly solved in very different ways. Some of the above mentioned tasks rely on proper motion segmentation provided by reliable background subtraction algorithm. Other methods are based on completely different approaches and do not use any kind of explicit segmentation. Such methods utilize classifier or bag of classifiers positioned at all the possible sub-windows in a given frame and identifies the location of various objects of our interest in the scene. We need to put some tracking algorithm on the top of the processing stack to track the moving objects.

During the development phase, we need to define some reliable evaluation methodology. One of the key aspects of real world surveillance application evaluation is robustness. Robustness is defined by the IEEE as the degree to which a system or component can function correctly in the presence of invalid inputs (e.g. data are deviated by the presence of

noise) or stressful environment conditions (e.g. phase of the day or various lighting) [78]. Human visual perception is very accurate and can be used to generate ground truth data that can be used later to judge the overall quality of designed algorithms and methods. Also, the choice of video sequence on which the algorithms are evaluated has a large influence on the results. Desired computer vision system should imitate the outcomes of human operator and we also expect that it will be able to guarantee stable results over a long period of time [106]. The humans, on the other side, are quite poor performers in the areas like monitoring crowded areas or fast moving objects. Hoverer, many researchers in computer vision are using different testing datasets making comparisons of resulting systems a bit difficult.

1.2 OUTLINE OF THE THESIS

The rest of the thesis is organized as follows. Chapter 2 brings the state of the art algorithms and methods used in the area of motion detection, tracking and vacant parking spaces detection. Author's contribution and the achieved results are given in Chapter 3. Partial improvement of well known histogram of oriented gradient is presented in Chapter 4. Chapter 5 brings the extension of the mathematical methods used for computation of confidence maps into the field of direct volume rendering. Conclusion and final thoughts are summarized in Chapter 6. List of author's publications is included in Section 7.4.

2

STATE OF THE ART

This chapter deals with the state of the art in the area of motion detection, object tracking and classification used in visual-based traffic surveillance systems. We will focus on classical approaches that were successfully applied in the above mentioned areas as well as in modern approaches improving the accuracy and reliability of surveillance systems. The attention will be paid especially to the algorithms and methods exploited in the application area of traffic density evaluation and parking spaces occupancy estimation.

The motion detection is the key component in the process of suspicious behaviour analysis. Hence, we will briefly review some of the techniques for background subtraction in Section 2.1. The basic principles of object recognition and classification, which are indispensable for semantic labeling and image understanding, are mentioned in Section 2.2. The object tracking aims at establishing the correspondence between objects across frames. These algorithms also have high relevance for every surveillance application, hence Section 2.3 is dedicated to this topic. Finally, the review of vacant parking space detection methods follows in Section 2.4.

2.1 MOTION DETECTION

The main goal of motion segmentation algorithms is to distinguish between moving foreground and stationary background. Frame differencing of temporally adjacent frames has been well studied since the late 70s [80]. The problem domain was originally constrained to segmenting vehicles from a roadway. A naive approach is to detect the foreground objects as the difference between the actual frame I_t (e.g. real function having a bounded support and a finite nonzero integral) and an image B representing the static background of analyzed scene. Two major inconveniences appear as soon as the image of static background is unavailable or the background is not strictly static at all. This analysis has led the researches to think of background image as the spatio-temporal function.

In fact, both proposed issues are very common and have to be treated by every successful background subtraction algorithm.

Let us state some of the most fundamental rules and constraints that must be obeyed. Firstly, the background image has not to be fixed. Secondly, the background model should also be adaptable to gradual or sudden illumination changes caused by various natural processes like Sun's movement, clouds, reflections, etc. Thirdly, the model should also accommodate to the permanent observed scene changes. And finally, camera oscillations and moving background objects such as trees or waves can harm the overall performance of background subtraction process too. Even small motions can cause large variations in the observed pixel values. These requirements can be met by building a representation of the scene called the background model and then we can find deviations from the model for each incoming frame [146]. We refer to this process as the background subtraction.

We have addressed the aforementioned problem related to camera oscillations and we will treat this in Chapter 4 devoted to the HOG features stabilization. But, in the following section, we will review some of the most common techniques for motion detection.

2.1.1 *Early Methods for Motion Detection*

The basic method is frame differencing (2.1) of temporally adjacent frames. The estimated background image is just the previous frame I_{t-1}. It evidently works well only in the case of particular conditions. The method is also very sensitive to the choice of threshold τ. The function m represents a foreground mask in which the value 1 stands for a foreground pixel and 0 means a background pixel

$$m(x,y) = \left\{ \begin{array}{ll} 1 & |I_t(x,y) - I_{t-1}(x,y)| > \tau \\ 0 & \text{otherwise} \end{array} \right. . \tag{2.1}$$

Ideally, the threshold τ should be a function of the spatial location (x,y). For example, the threshold should be smaller for the regions with low contrast. One possible modification is proposed by Fuentes and Ve-

lastin [66]. They use the relative difference (2.2) rather than absolute difference (2.1) to emphasize the contrast in the dark areas

$$m(x,y) = \begin{cases} 1 & \frac{|I_t(x,y) - B_t(x,y)|}{B_t(x,y)} > \tau \\ 0 & \text{otherwise} \end{cases} .$$

(2.2)

Velastin [99] and Cuchiara [43] define the background B_t as the average, resp. the median of previous n frames. However, such methods are very memory consuming. Chien et al. [145] surmise that if a pixel is stationary for the past several frames, then the probability is high that it belongs to the background region. A background registration technique is used to construct a reliable background image from the accumulated frame difference information. They also show the importance of shadow handling.

Elgammal et al. [53] present a non-parametric background model and a background subtraction approach. The probability density function of the pixel intensities is estimated using the normal kernel estimator K as

$$P_{x,y}(I_t(x,y) = u) = \frac{1}{L} \sum_{i=t-L}^{t-1} K(u - I_i(x,y)).$$

(2.3)

During the subtraction process, the current pixel is matched not only to the corresponding pixel in the background model, but also to the neighborhood locations. Thus, this model can handle situations where the background of the scene is cluttered and not completely static but contains small motions such as tree branches and bushes.

Cheung and Kamath [42] propose the survey of existing background subtraction algorithms supplemented by experimental results. They discriminate background modeling as the heart of any background subtraction algorithm into two categories. The first one, denoted as non-recursive techniques, uses a sliding window approach for background estimation based on temporal variation of each pixel. Such techniques are highly adaptive and require a large frame buffer. Some of the earlier mentioned techniques including frame differencing, median filter and linear predictive filter belong to this category. The second one, referred as recursive techniques, do not maintain a buffer for background estimation. They recursively update only a single background model derived from the stream of incoming frames. As a result, all preceding frames affect the current background model. Most schemes incorporate exponential decay

to reduce the influence of very old frames on actual model and make it more flexible. The collection of the most remarkable algorithms that belong into this category include approximated median filter [105, 119], Kalman filter [90] and Mixture of Gaussians (MoG) [65, 115].

The work of Friedman [65] explicitly models the values of all the pixels as one particular type and classify the pixel values into three predetermined distributions. The simple form incorporating only one Gaussian distribution proposed in [139] is not sufficient to represent multimodal background. Multimodality is a very natural property of nearly every real image sequence and occurs in situations when some pixel is periodically occupied by a small closed set of brightness values. Such behaviour is often related with the areas containing high frequencies (e.g. tree branches, water level).

Especially, MoG enjoys tremendous popularity because of its ability to effectively model multimodal background. Stauffer and Grimson [127] model the values of a particular pixel as a mixture of Gaussians

$$f_{\mathbf{X}|k}(X|k,\Theta_k) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_k)^\mathsf{T}\Sigma_k^{-1}(X-\mu_k)} \, , \qquad (2.4)$$

where $\Theta_k = \{\mu_k, \sigma_k\}$ and the covariance matrix $\Sigma_k = \sigma_k^2 \mathbf{I}$. The index $\hat{k}$ denotes the particular surface appeared on the position of a modeled pixel and can be obtained with the aim of the Bayes theorem as follows

$$\hat{k} = \arg\max_k P(k) f_{\mathbf{X}|k}(X|k,\Theta_k) \, . \qquad (2.5)$$

By the 1999, when MoG appeared for the first time, many reimplementations were created. The tutorial paper [115] describes a practical implementation of the Stauffer-Grimson algorithm and provides adequate information about the underlying statistical theory. In the report [30], Bilmes studies the maximum-likelihood parameter estimation problem and how the Expectation Maximization (EM) algorithm can be used for its solution.

As can be seen in Fig. 2.1, a generic background subtraction algorithm consist of few common parts. The most important one is depicted in the flow diagram as a background modeling process. But for a good overall tracking performance, the preprocessing and data validation steps are indispensable.
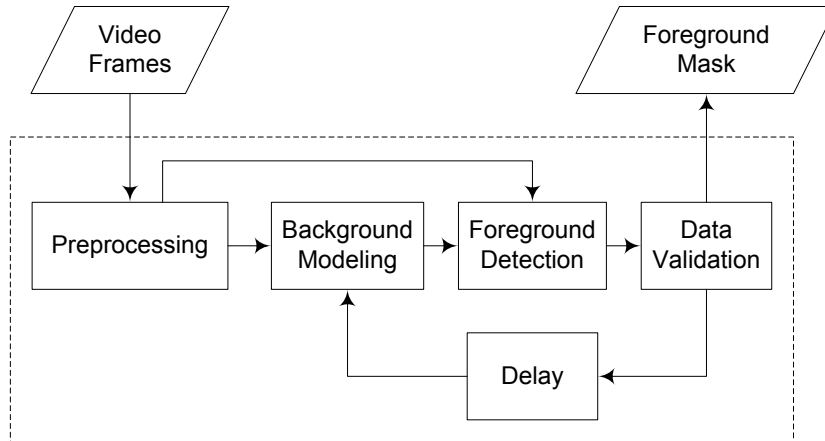
Figure 2.1: Flow diagram of a generic background subtraction algorithm [42]

Another approach how to capture the motion in the image sequences is optical flow. Optical flow is a dense field of displacement vectors that defines the translation of each pixel [146]. It is computed using the brightness constraint, which assumes brightness constancy of corresponding pixels in consecutive frames [69]. However, this definition gives rise to the ill-posed problem. This is the well known aperture problem in optical flow (see Fig. 3.3). A popular method for computing optical flow is the differential method of Lucas and Kanade [101] solving the basic optical flow equations by the least squares method. Horn and Schunck [69] combined the gradient constraint with a global smoothness term to estimate the velocity field **v** minimizing the expression

$$\iint_{\Omega} (\nabla I \cdot \mathbf{v} + I_t)^2 + \lambda^2 (\|\mathbf{v}_x\|^2 + \|\mathbf{v}_y\|^2) \, \mathrm{d}x \, \mathrm{d}y \qquad (2.6)$$

over the image domain Ω, where λ controls the influence of the smoothness term. In order to handle discontinuities in optical flow, many new concepts were developed. In [36], the authors proposed a novel variational approach that integrates coarse-to-fine strategy using the so-called warping technique with a solid numerical method.

We should also take in account the influence of environment in which the camera is placed. Unfortunately, complications occur on a regular basis. One of the most apparent problem in motion detection is an unstable camera support. In [82], the authors provide a method for reducing cam-

9

(a) Rotating barber's pole   (b) Real motion field   (c) Optical flow field
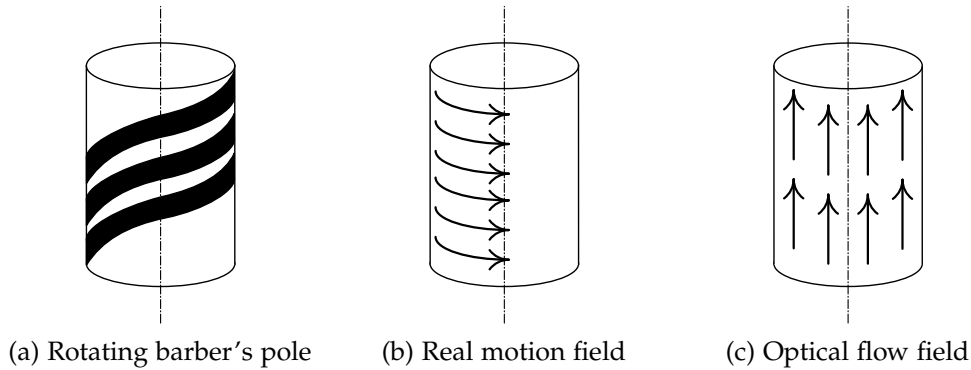
Figure 2.2: (a) Barber's pole is rotating around its length-axis and induces apparent motion of black strips pattern in the image. Observer is unable to judge the correct motion field (b) as we are able to measure only the component of the optical flow (c) that is in the direction of the intensity gradient. This phenomenon is known as an aperture problem

era jitter. An object is assumed to be moving if its dynamical behaviour is different from the average dynamics observed in a reference sequence.

Im et al. in [79] propose wavelet-based moving object segmentation using high frequency coefficients in wavelet subbands. Antić et al. [25] propose a novel wavelet-based method for robust feature extraction and tracking. They also claim that extremely harsh conditions can occur and violate the premises of the statistical regularity and predictability of background pixels. The goal of our paper [2] was to eliminate the unfavourable influence of camera related phenomena like sudden changes of overall image brightness caused by the auto-exposure control and to overcome camera jitter from unstable support.

2.2 OBJECT RECOGNITION AND CLASSIFICATION

In the previous section, we have mentioned a few common methods specialized on the detection of motion in image sequences. To some extent, we considered this process as a classification of independent pixels in time. The global differential methods for optical flow computation are an exception. In this section, we will review some higher-level classification methods that are fundamental for image understanding.

Associating the regions in images with semantic labels can significantly improve the overall performance of many computer vision applications. The main objective is to group the individual pixels into clusters based on low-level features. These features are crucial for image representation and should preserve all relevant information of the input imagery in a compact way. The amount of information in the feature set or feature vector is also constrained by the learning model. Early works used the features extracted regularly from the whole image. More recent approaches exploit the strategy of patches extracted at the points of interest. The representation can be based only on appearance or the location can be included. The invariance of features is an another important factor. We recognize the invariance to rotation, scale, translation, illumination, occlusion, and clutter. From the statistical viewpoint, object categorization can be divided into discriminative methods modeling posterior probabilities (i.e. $P(\text{class}|\text{image})$) and generative (or descriptive) methods modeling likelihood and prior (i.e. $P(\text{image}|\text{class})$ and $P(\text{class})$)

$$\frac{P(\text{class}|\text{image})}{P(\neg\text{class}|\text{image})} = \frac{P(\text{image}|\text{class})}{P(\text{image}|\neg\text{class})} \frac{P(\text{class})}{P(\neg\text{class})}. \qquad (2.7)$$

For example, MoG and other types of mixture models belong to generative methods. In the case of generative methods training, we are maximizing the likelihood. Discriminative models include logistic regression, support vector machines (SVM), boosting, neural networks. For the training of discriminative methods, we need a set of negative images. Object classification is also significantly affected by semantic context in image [118, 117]. Generally speaking, contextual recognition is maximizing a scene probability function that incorporates the output of individual object detectors and pairwise interactions between the objects [88].

2.3 OBJECT TRACKING

The aim of object tracking is to generate the trajectory of a moving object over a period of time [146]. The main function of object tracker is to establish the correspondence between the successive instances of tracked object across the individual frames. Object tracking is still a non trivial problem by virtue of the projection of 3D real-world coordinates into 2D

image plane, complex motion and geometry of nonrigid objects, partial and full object occlusions.

Tracking can be formulated as frame by frame association of detections based on geometry and dynamics without particular appearance models. Other class of trackers utilizes the appearance models coupled with geometry and dynamics. The most advanced approaches furthermore integrate detection and tracking in a Bayesian framework, combining the appearance models with an observation density, dynamics, and probabilistic inference of the posterior state density [56]. The shape and appearance models significantly affect the motion models. For example, if an object is represented as a single point, then only a simple translational model can be used [146]. Most of the rigid objects can be represented by regular geometric shapes like an ellipse. The shape of nonrigid objects may be characterized by a silhouette or contour.

The general problem of every object tracking algorithm is how to deal with multiple targets in the image. One possibility is to generate a joint-space involving the number of targets and their possible configurations. Unfortunately, the dimension of state space grows rapidly. In this case, the computational complexity can be reduced by sophisticated sampling techniques like Metropolis-Hastings [91]. Another commonly used solution is to assign one tracker per each object, which simplifies the state space representation [89]. Recently, Kalal et al. [87] present interesting framework combining median-flow tracker, P-N learning and on-line trained 1-NN classifier.

2.4 STATIC TRAFFIC SURVEILLANCE

In this section, we will focus on a specific kind of traffic surveillance systems. So called static transport represents the transport branch specialized on stationary aspects of transportation. In the present time, we are the witnesses of rapidly deteriorating situation in car parking. It is not just a waste of time for the driver to find a free parking space, it also has a negative impact on pollution in large agglomerations. From our point of view, the problem of identifying free parking spaces in a large parking lot is a quite interesting task. Despite the fact that the parked cars are mostly static, there is a lot in common with moving cars surveillance.

In the past decade, quite a lot of works concerned in vacant parking space detection appeared. There exist four main categories of parking guidance systems using different technologies including the counter-based, wired-sensor-based, wireless-sensor-based and vision-based approaches [31]. The vision-based systems promise a number of advantages over the intrusive sensors, like inductive loops, infra red sensors, a combination of ultrasonic and magnetic sensors [96] and other detectors mentioned in Chapter 1. Image analysis provides several effective techniques to detect the presence of various objects captured within camera's field of view. The first parking guidance information system was deployed in Aachen, Germany in 1971 [98]. The goals of parking lot surveillance include counting parked cars, identifying the location, size or type of parked vehicles, monitoring the movement of cars and activities of humans. In the following text, we will try to give an overview of the state of the art in parking space detection.

The parking space state is fully described by only two disjoint labels indicating the place as occupied or vacant. These two states are the elements of the output set $\mathcal{L} = \{o \equiv \text{occupied}, v \equiv \text{vacant}\}$. From this point of view, the solution seems to be easy. In reality, we are facing to various conditions significantly influencing the whole process of parking lot images analysis. Our main goal is to find a function that will be, under certain conditions, able to assign a correct state from the set $\mathcal{L}$ to each parking space with high confidence.

The previously published methods can be divided into three distinct classes that differ in the principle used to determine the parking space state. The first class of those methods rely on a proper car detector. When no car is overlapping the area of parking space, the space is marked as vacant. To detect the location of the object of interest in image we can use various visual object recognition methods, such as comparison with reference images using normalized principal component of feature characteristics, path tracking and other methods mentioned in Subsection 2.2.

In [137], Wang and Hanson introduced the parking lot analysis from aerial images based on extraction of the height features of objects on the ground with a stereo terrain reconstruction algorithm. The elevation map is generated from a pair of images and all the elevated areas correspond with parked vehicles, while the ground areas appear to be flat. The surface inclination can violate simple global thresholding algorithm so the authors have solved this problem by the surface texture and microstruc-

ture extraction system where the microstructures (sizes are near the limit of image resolution) are distinguished from large scale structures (buildings). The extracted rectangles may cover more than one vehicle. To separate the adjacent cars, the author suggest to utilize the a priori knowledge of typical car size and the parking spaces markers that can be considered as regularly aligned microstructure pattern. Usage of this approach is rather limited for sparsely populated parking lots. This limitation can be reduced by combination with Chellappa's et al. system [40]. This approach is more relevant for a specific use than parking vacancy monitoring. More recent paper of Yang et al. [143] introduced the fast cascade boosting approach for detecting vehicles in 2D grayscale aerial images.

Yamada and Mizuno [142] analyzed the distribution of segments in parking cell and their work has significantly influenced other researchers. The level of cell fragmentation is supposed to be related with the presence of car. Vehicles are composed of numerous components that are usually rendered as segments and the number of segments, weighted by area, is used to decide whether a place is vacant or occupied. The following score $g = (\sum_{k=1}^{n} k S_k)/(\sum_{k=1}^{n} S_k)$ is used to categorize the state of parking space. Here n denotes the total number of segments, and $S_1, \ldots, S_n$ denote the areas of individual segments in decreasing order. The score g is higher if the segments are more uniform. Only a single threshold is used to evaluate the actual state of every monitored parking place. Contrary to [137, 40], common camera placed at a high position is used instead of aerial images. The authors reported a 98.7 % successful detection rate.

Funck et al. [67] assume that the constant video streams are not available in common surveillance installations. Therefore, the techniques like motion detection and object tracking are not applicable and the entire process of parking lot analysis rely only on a single image. They also suggest that the detection of individual vehicles is not promising. Instead, the parking place status is estimated comparing the whole image to the reference image of empty places. The reference base should contain all possible illumination situations, but this is not feasible. So the authors build a model of the empty place by applying PCA on a number of images under different illuminations. Via eigenspace reconstruction, they reconstruct the image of empty parking lot under the present illumination conditions and all cars will disappear. The difference image is further thresholded and morphologically filtered. The authors also admit that the texture properties of vehicles are very similar to those of the parking

spaces, which limits the capabilities of texture analysis to the cases where the cars are distant enough. Testing the system on real world sequences yields an average error rate of approximately 10 %.

Foresti et al. [63] focused on the high-level event recognition to analyze the events observed in the scene of parking lot. The moving objects are extracted by comparing the input image with a background image and the background model is updated using the Kalman filter. Blob tracking is based on the Mean Shift algorithm. Object classification is done by adaptive high order neural tree (AHNT) learned by feature vector composed of eight distances representing the geometric properties of each blob. The preliminary object classification obtained on the basis of n consecutive frames is analyzed in a winner-take-all fashion. The proposed system is also able to detect suspicious behaviour of pedestrians. A typical trajectory of walking pedestrian is almost rectilinear for a long time. An off-line event database is built of features containing the coefficients of Bézier curve that approximates the object trajectory. Spatially and temporally consecutive simple events are further hierarchized creating composite events. The composite events represent a complex chain of successive actions, like moving vehicle stopping in a given parking space, a person exiting that vehicle and moving out of the parking lot area. Any violation of the predefined rules may cause an alarm.

Lin et al. [98] use an adaptive background model for each parking space. They suppose that the distribution of parking spaces is known in advance. To prevent the color from shifting due to different weather conditions, a color balancing algorithm, more accurately gray world assumption, is used. The background images are obtained with the aim of median applied on the sequence of 60 images. The background pixels of every parking place are represented by the mean and standard deviation. The foreground is extracted based on the difference of the background image and the actual frame. The influence of shadows is deduced from the foreground images using chrominance information.

Wu and Zhang [140] introduced a new approach in the sense of shadow and occlusion handling. Wu et al. also published, in a slightly improved version, the same approach in [141]. Each parking row is extracted from the parking lot image. The parking spaces are grouped into detection patches containing three adjoining parking spaces. During the training process, the authors manually classify all parking spaces with eight space statuses encoding the state of each place in the patch. The ground color

model is represented by a single Gaussian representing the probability of the pixel value belonging to the parking lot background. Only the three scanning lines are used to get the values of pixels from the normalized image of patch. In this way, 75×3 features per patch are extracted. PCA reduces the number of critical features to 50. The classification is done by the general binary SVM classifier adapted by one-against-one strategy that takes all possible two-class combinations. Therefore, 28 SVMs are trained. Different labeling of overlapping patches can occur and, basically, there are three kinds of conflict states: no conflict, one conflict, two conflicts. The authors applied Markov random field (MRF) to correct this labeling inconsistency.

In [132], True manually extracts the overhead images of individual parking spaces. Then the author converts obtained images from RGB color space to L*a*b and throws away the luminance channel. Afterwards, the color histograms for both chrominance channels are constructed. Each histogram contains 32 bins and is classified by k-nearest neighbour classifier equipped with Chi-squared distance χ^2. Alternatively, the author also used a binary SVM to classify the parking spaces based on the assumption that all vacant parking spaces have a similar color. The Harris corner detector is used to extract significant features that are in a vocabulary of vehicle features and another of non-vehicle features. Vocabulary consists of small 25×25 images centered at each feature point. The individual features found in the parking space image are compared to the features in the vocabulary using normalized cross correlation for final classification. Unfortunately, the author does not provide a method for combining the results of color histogram classification and vehicle feature detection. Classification accuracy varies between 69 % and 91 %.

Sastre et al. [121] propose a methodology to compute a 2D homography applying the homomorphic method for solving systems of equations over $\mathbb{Z}$. The Gabor filters are proposed for texture feature extraction of the background of parking lot since the Gabor filters can be considered as an edge detector. A two dimensional Gabor function $g(x, y)$ and its Fourier transform $G(x, y) = \mathcal{F}\{g(x, y)\}$

$$G(x, y) = \exp\left(-\frac{1}{2}\left[\left(\frac{u - W}{\sigma_u}\right)^2 + \left(\frac{v}{\sigma_v}\right)^2\right]\right), \qquad (2.8)$$

where $\sigma_u = 1/(2\pi\sigma_x)$ and $\sigma_v = 1/(2\pi\sigma_y)$. The Gaussian's major and minor axis are determined through σ_x and σ_y. A Gabor feature image

16

$S(x,y)$ is obtained by convolution of an input image $I(x,y)$ with the Gabor function $g(x,y)$ as follows

$$S(x,y) = \iint\limits_{\Omega} I(\phi,\varphi)g(x-\phi,y-\varphi)\,\mathrm{d}\phi\,\mathrm{d}\varphi.$$ (2.9)

To extract the texture features, the authors introduced a generating function enabling scaling and rotation of the mother Gabor function $g(x,y)$. Generated dictionary of self-similar Gabor filters forms a complete non-orthogonal basis set implying the presence of redundant information in the filtered image.

Huang et al. [71, 70] proposed a robust parking space detection based on the three-layer Bayesian hierarchical framework (BHF). The detection framework consists of an observation layer, where each node represents a local feature, the middle layer is a hidden labeling layer representing the categorization of local region and the third layer indicates the semantic hypotheses of the parking spaces. The inter-layer connection is subject of actual parking spaces adjacency. With the aim of Bayesian statistics, the parking space detection problem can be solved by determining of a maximum a posteriori probability (MAP). The authors aim to find the most reasonable space status S_L and pixel labeling H_L based on the image partition D_L according to the equation

$$H_L^*, S_L^* = \arg\max_{H_L, S_L} P(H_L, S_L | D_L).$$ (2.10)

Nallamuthu and Lokala [110] compute, for every pixel in the actual image, the absolute difference with the reference image of the empty parking lot. The individual pixels are considered occluded or unoccluded according to the resulting values and an experimentally estimated threshold. The interocclusions are handled through manually marked regions that get occluded only when there is a car parked in particular space. The ratio of occluded to non occluded pixels indicates the parking space state. The secondary approach is based on the difference of the 3D color histograms of the background and the current image. The euclidean distance between these two histograms identifies the parking state as well. In the third approach, the authors use the texture matching and the morphological operators to distinguish between the parking space states. Reported success rates vary between 59 % and 70 %.

Bong et al. [31] begin with automatic identification of every parking spaces in the image. Described approach is restricted to the situation when white lines separating the individual parking spaces are visible during the initialization phase. An unoccupied parking space is represented only by a single value and actual state is estimated via thresholding. To minimize the influence of projected shadows the authors suggest the use of median filtering and Sobel operators for edge detection. The final occupancy detection accuracy is more than 93 %.

Ichihashi et al. [75, 76] introduced a parking space state classifier based on fuzzy c-means (FCM) clustering and hyper parameter tuning by particle swarm optimization (PSO). FCM is an extension of k-means, the popular simple clustering technique [74, 73]. While k-means discovers hard clusters (a point belong to only one cluster), FCM is a more statistically formalized method and discovers soft clusters where a particular point can belong to more than one cluster with certain probability.

Chen et al. [41] proposed a system with multiple cameras for monitoring a wide parking area. Affine transformations are used to merge the images obtained from these cameras. The system consists of two components. Object tracking is based on background subtraction and connected component analysis. Parking space detection is carried out by an edge based scheme. The authors assume that, when a vacant parking space is occupied by a vehicle, its edge density will increase. To tackle the occlusion problem, the parking space is divided into four weighted cells.

The method of Dalka et al. [46, 47] works similarly to [41]. The moving objects are detected by background subtraction utilizing MoG and the resulting blobs are tracked with Kalman filter. To reduce the false-positive results, only the objects entering the camera field of view from the outside are taken into account.

Some specific type of parking lot management system is represented by automatic parking systems. These systems extend the variety of on board systems of modern cars and can be regarded as an evolution of common parking sensors. A vision-based parking assistant systems for autonomous parking [86] is based on L-shaped pattern search in the depth map. This first estimate is refined by a more sophisticated generic vehicle surface model minimizing the distance between the object model and the point cloud.

Jung et al. [83] propose stereo vision-based localization of free parking site for automatic parking system. They use the four-pixel classes

reflecting the intensity configuration of pixel neighbourhood as features for stereo matching which is fast and robust to noise. Stereo matching is performed only on pixels classified as vertical edge. The proposed system was later extended in [84, 85].

Suhr et al. [131] developed a vacant parking space detection system using motion stereo based 3D reconstruction. An image sequence is acquired with a single rear view fish eye camera and the view behind the automobile is three dimensionally reconstructed by using point correspondences. The detection accuracy of roughly 90 % was achieved.

Bravo et al. [34] considered two kinds of features for the parking space analysis: edge density [41] and pyramidal histogram of oriented gradients (PHOG) proposed by Bosch et al. [32]. PHOG features represent an image by its local shape and the spatial layout of the shape. To classify every parking space, its PHOG features are concatenated and compared with training samples of occupied parking spaces using a Chi-squared distance χ^2. On the three 6-hour test sequences, they achieved the accuracy ranging from 0.821 to 0.937 for PHOG and 0.523 to 0.853 for the edge density features.

Recently, Suhr and Jung [130] presented a vacant parking slot detection and tracking based on around view monitor (AVM) system coupled with an ultrasonic sensor-based automatic parking system. Parking slot marking in AVM image sequences are detected with the hierarchical tree structure-based method [129] while the vehicle is passing by.

It is important to note that some of described approaches used images with very minimal vehicle to vehicle occlusion [132]. Ablavsky et al. [24, 23] proposed a representation for scenes containing relocatable objects that can cause partial occlusions. Also, many test sequences are biased towards ideal weather conditions with the exception of [75, 76].

Seo et al. [125, 124, 123] address the general problem of extracting the entire parking lot structure from overhead imagery. Their approach is based on straight lines extraction. The authors suggest to compute the intensity gradient at each pixel and then quantize the gradient into the directions using predefined ranges. A connected component algorithm is then used to group the pixels assigned the same direction to form the line supporting regions. The first principal eigenvector of a line supporting region determines the direction of the line. The lines that are either too short or too long from each of the line clusters are removed. The remain-

ing lines are used for estimating the parameters of a parking row. SVM is used to filter out the false parking spots.

2.5 A REVIEW OF THE CURRENT STATE

Visual-based parking guidance systems are experiencing increased interest from parking operators, local authorities and also car manufacturers. We can expect that this will be a long-term interest. From the viewpoint of computer vision community, there are a lot of open questions. The most recent publications introduced three-layer Bayesian hierarchical framework [70] and the fuzzy c-means clustering [75] into the parking space status decision process and at this moment, these two works can be considered as the most advanced methods in this area. However, the authors do not reflect the influence of, e.g. pedestrians on the status inference. The performance of vacant parking space detection algorithms deteriorate in the case of pedestrian presence. Also the influence of the unpredictable lighting changes caused by car headlights is not solved in any of the referenced paper. There is also lack of utilization of a priori information about the parked vehicles geometry. Another problem is connected with inter-object occlusions. Inter-space correlation proposed in [141] significantly increased the detection accuracy via Markov Random Field based correction.

3

PARKING SPACES OCCUPANCY DETECTION

In this chapter, we present an algorithm for estimating the occupancy of individual parking spaces. Our method is based on a computer analysis of images obtained by a camera system monitoring the activities on a parking lot. The proposed method extensively uses a priori information about the parking lot layout and the general shape of well-parked cars, which is incorporated in a simplified probabilistic car model. Discriminative features are extracted from a normalized image of every parking space, the relevance of these gradient-based features is prioritized via a selective flow, and furthermore, their spatial relationship is revealed through an undirected graphical model. In contrary to other methods, we strive to avoid the training phase to reduce the time required to bring the system into a fully operational state. The reliability of the here devised approach is evaluated on the set of video sequences captured during different phases of a day. The results are compared against the ground truth data and the most profound methods as well.

CITATION   The method presented here was published in [1, 4, 3] and is cited in 17 conference papers, journal articles and dissertation thesis among them in [58, 72, 122, 135, 81, 136, 50, 103].

3.1 BRIEF OVERVIEW

Video-based surveillance systems had evolved into sophisticated systems during the last 70 years. During that time, CCTV systems had successfully spread through many various application areas including monitoring dangerous industrial processes, security systems in banks, streets, stores, systems supporting transport safety and traffic surveillance. In this chapter, we will focus on a specific kind of traffic surveillance systems, on the so-called parking lot guidance systems and the related area of image analysis as the vision-based systems promise a number of advantages over the intrusive sensors [96]. The very first parking guidance
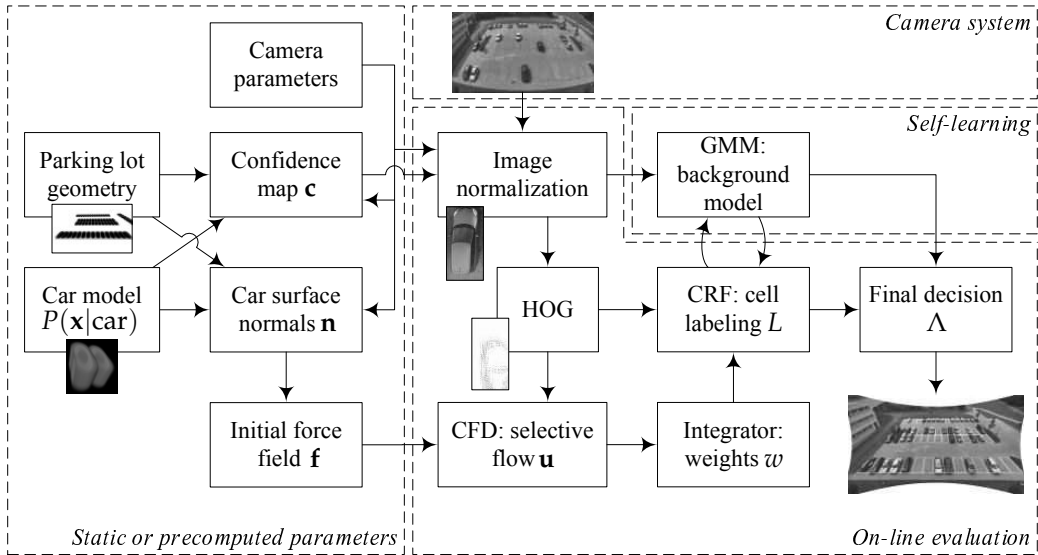
Figure 3.1: The overview of our parking surveillance system. A priori known parking lot geometry, camera parameters, and probabilistic car model provide all the data necessary to calculate the confidence map for occlusion handling and expected car surface normals which form the force field initiating the advection of HOG features obtained from normalized images. The resulting flow prioritizes features that are conformal to the expected shape of a car and assign a weight to each cell. The final decision about the parking space state is devised from the labeling produced by CRF minimizing the related Gibbs energy consisting of potentials based on weight values and background GMM of an empty parking space

information system was deployed in Aachen, Germany in 1971 [98]. In the present time, the problem of identifying free parking spaces in a large parking lot is a quite interesting task. In the past decade, quite a lot of work concerned in the vacant parking space detection appeared, e.g. [141, 70, 76]. There exist four main categories of parking guidance systems using different technologies including the counter-based, wired-sensor-based, wireless-sensor-based and vision-based approaches [31]. The goals of parking lot surveillance include counting parked cars, identifying the location, size or type of parked vehicles, monitoring the movement of cars and the activities of humans.

3.2 SHADOWS ATTENUATION

Shadow removal can be identified as a critical step in general processes like object detection and tracking. Related methods can be categorized according the feature based taxonomy into four categories: chromacity, physical, geometry, and textures [120]. Here we focus briefly on creating illumination invariant image from a single image acquired by common digital still camera. These invariant images are also referred as intrinsic images to express their independence of lighting. Only the intrinsic reflectivity of the object is captured and therefore object look like it has no shading. Every single pixel of image is made up of two components. Illumination characterizes light source (illuminant) and reflectance, property of reflecting surface. Color components R_k, $k \in \{R, G, B\}$ of every pixel **R** in RGB color space can be described by equation

$$R_k = \sigma \int_\omega E(\lambda)S(\lambda)Q_k(\lambda)\,\mathrm{d}\lambda, \tag{3.1}$$

where σ denotes Lambertian shading, E represents illumination spectral power distribution function, S is spectral reflectance function and Q_k is for camera sensor sensitivity [52]. The integral is taken over visible wavelengths of light ω. If camera sensors are narrow band (i.e. similar to Dirac's delta function) so that $Q_k(\lambda) = q_k\delta(\lambda - \lambda_k)$, then we can reduce Equation (3.1) to the form

$$R_k = \sigma \int_\omega E(\lambda)S(\lambda)q_k\,\mathrm{d}\lambda, \tag{3.2}$$

with $q_k = Q_k(\lambda_k)$ [60]. Furthermore, illumination function E can be approximated by Planck's law modified for typical temperature ranges of most light sources

$$E(\lambda, T) = Ic_1\lambda^{-5}e^{-\frac{c_2}{T\lambda}}, \tag{3.3}$$

where c_1 and c_2 are constants, I controls light intensity and T is the temperature of light [60]. Substituting term $E(\lambda, T)$ in Equation (3.2) by Equation (3.3), we get narrow-band sensor response

$$R_k = \sigma Ic_1\lambda_k^{-5}e^{-\frac{c_2}{T\lambda}}S(\lambda_k)q_k. \tag{3.4}$$
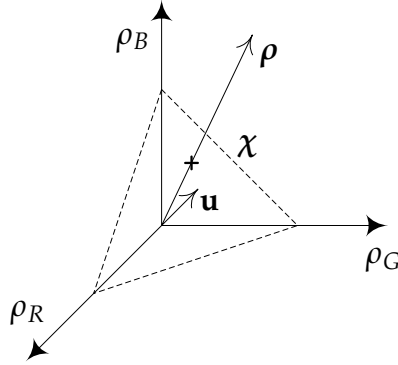
Figure 3.2: An example of projection of vector $\boldsymbol{\rho}$ onto a 2D chromaticity plane (*dashed triangle*)

If every color component R_k is divided by it's geometric mean R_M, we will remove intensity information I and Lambertian surface σ [52]

$$c_k = \frac{R_k}{\sqrt[3]{\prod_{i \in \{R,G,B\}} R_i}} = \frac{R_k}{R_M}. \tag{3.5}$$

Now, the log version of c_k is taken

$$\rho_k = \log(c_k). \tag{3.6}$$

For each color pixel R we get a vector $\boldsymbol{\rho} = (\rho_R, \rho_G, \rho_B)^\top$ representing a point which lies on a plane orthogonal to the vector $\mathbf{u} = 1/\sqrt{3}(1,1,1)^\top$. We can project all these points onto a 2D chromaticity space $[\chi_1, \chi_2]$

$$\begin{aligned} \mathbf{P}_u^\perp &= \mathbf{I} - \mathbf{u}\mathbf{u}^\top = \mathbf{U}^\top, \\ \chi &= \mathbf{U}\boldsymbol{\rho}. \end{aligned} \tag{3.7}$$

where 3×2 orthogonal matrix $\mathbf{U}$ is defined in [62] as

$$\mathbf{U} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}, \tag{3.8}$$

and the projector $\mathbf{P}_u^\top$ has two non-zero eigenvalues [61]. Described situation is for clarity depicted in Fig. 3.2.

The values (χ_1, χ_2) across different lighting tend to fall on a set of straight lines in the 2D scatter plot (Fig. 3.3b). All lines should be parallel
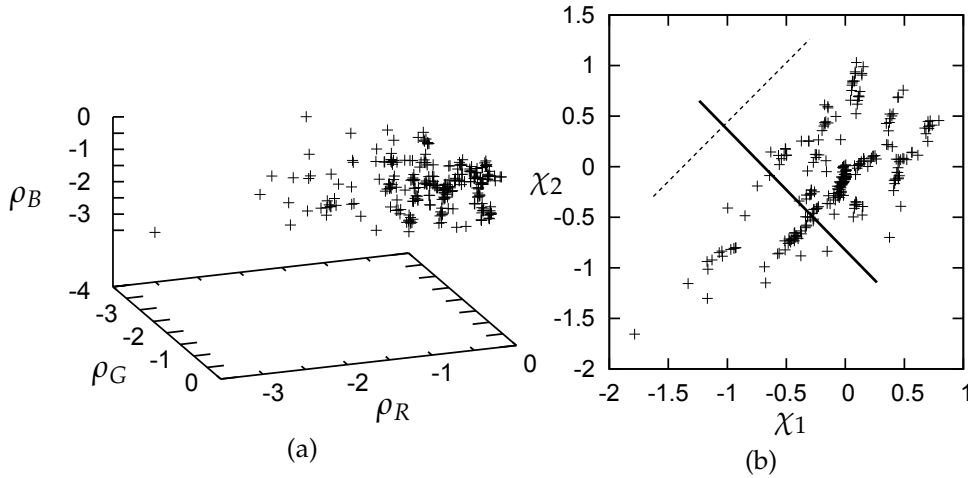
Figure 3.3: Calibration results: (a) **R** vectors of color patches from Gretag Macbeth color checker. (b) Same vectors as projected onto a 2D chromaticity space. Samples of same color patch captured under different lighting tend to lie on straight lines, e.g. *dashed line*. Projecting these samples onto the *thick black line* we obtain the shadow-free image

for given camera. As a final step, we get intrinsic image $\hat{I}$ with aid of the following equation

$$\hat{I} = \chi_1 \cos(\phi) + \chi_2 \sin(\phi).$$ (3.9)

To do so, we must find proper value of ϕ in which to project vector χ such that the effect of illumination is significantly attenuated or completely removed. This parameter depends on the camera being used. The simplest way to calibrate camera is to take pictures of multi-color surface under many different lighting conditions (e.g. during the whole day). In our experiments, we use Gretag Macbeth color checker. The main idea behind the process of finding angle ϕ is described in [61].

After evaluating all captured calibration images of color checker under different Planckian lighting, we get a set of ρ vectors (Fig. 3.3a). One can reintegrate shadowless intrinsic image back into the original color image, but this is not our case. Gray-scale image is fully suitable for the next step. Our tests confirm conclusions from [64], because there

25

is the obvious need for narrow band sensor. Sharpening procedure for commonly used cameras is required.

3.3 OCCLUSION HANDLING

Occlusions significantly affect the performance of object recognition and tracking algorithms. A lot of effort has been done in the area of occlusion handling in dynamic scenes, e.g. [111, 144]. Occlusions are also very common in parking lot images due to the spatial arrangement of parked cars and camera position and also some parking spaces may be heavily occluded by neighboring parked cars. In order to cope with inter-vehicle occlusions, we propose a probabilistic 3D model of a vehicle. This model represents all feasible positions of vehicle inside the parking space. In the most simplistic way, the model can be represented by a cuboid positioned at the parking lot surface. The model itself is fully defined by its width, length, height, position of center and yaw. These parameters are treated as independent normally distributed random variables. As a result, we obtain a 3D scalar field of $128 \times 128 \times 256$ values representing the probabilities that the particular region inside the volume over a single parking space belongs to a vehicle. This can be expressed as the likelihood $P(\mathbf{x}|\text{vehicle})$, where $\mathbf{x}$ represents some discrete element (voxel) inside this scalar field. To put this model in the relation with the camera, we can cast a ray through the continuous scalar field $\rho : \mathbb{R}^3 \to \langle 0, 1 \rangle$ which is obtained as a trilinear interpolation of the discrete fields of likelihoods $P(\mathbf{x}|\text{vehicle})$. With the aim of basic calculus, we can formulate the expression for the scalar field of occlusions h in terms of the line integral along a piecewise smooth curve L (line of view made up of the set of straight segments intersecting affected cells C, see for reference Fig. 5.1 on page 64) as follows

$$
h(x,y) = \bigcup_{\mathbf{x} \in C} P(\mathbf{x}|\text{vehicle}) = \int_L \rho(s) \, \mathrm{d}s = \int_a^b \rho(\mathbf{r}(t)) \, ||\mathbf{r}'(t)|| \, \mathrm{d}t
$$

$$
= \int_a^b \rho(\mathbf{r}(t)) ||\hat{\mathbf{d}}|| \, \mathrm{d}t = \int_a^b \rho(\mathbf{r}(t)) \, \mathrm{d}t = \cdots = \sum_{i \in I} \int_{t_i}^{t_{i+1}} \sum_{j=0}^{3} a_j t^j \, \mathrm{d}t, \quad (3.10)
$$

where the ray $\mathbf{r}(t) = O + \hat{\mathbf{d}} \, t$ is a bijective parameterization of the line segment originating at the point $\mathbf{r}(a)$ coincident with the camera's origin

O and the end point $\mathbf{r}(b)$ that is the intersection with the parking lot plane. In addition, the integral over the interval $\langle a, b \rangle$ is decomposed into the sum of integrals over the set *I* of intervals $\langle t_i, t_{i+1} \rangle$ representing the parametric coordinates of intersections of the ray $\mathbf{r}$ with the set of affected cells *C*. The analytical derivation of parameters a_0, a_1, a_2 and a_3 is devised in Chapter 5 and we left them out of calculation in this chapter to avoid unnecessary clutter. Returned scalar value of the function *h* represents the degree of our believe that the certain position (x, y) in the image of parking lot surface can be occluded exclusively by a well-parked car (see Fig. 3.4a). The resulting confidence field for the *i*-th parking space equals to $c_i(x, y) = 2h_i(x, y) - h(x, y)$, where h_i is the occlusion map where only *i*-th parking place is occupied and *h* is the occlusion map generated for the fully engaged parking lot. Figure 3.4b presents the unprojected confidence field for the first parking space and the projected version of the same field is in Fig. 3.4c.

3.4 PARKING SPACES EXTRACTION

To extract the normalized images of individual parking spaces (see Fig. 3.5), we use a so-called pinhole camera model

$$\mathbf{p}' = \mathbf{A} \left[\mathbf{R} | \mathbf{t} \right] \mathbf{p} \tag{3.11}$$

projecting a 3D point $\mathbf{p}$ in homogeneous coordinates from the world-space into the image plane using a perspective transformation. The joint rotation-translation matrix or matrix of extrinsic parameters $[\mathbf{R}|\mathbf{t}]$ is revealed from a priori known correspondence between the set of at least four object points and image points. The algorithm applied here is devised in [147]. Furthermore, a camera matrix $\mathbf{A}$ contains the camera intrinsic parameters like principal point and focal lengths. The wide-angle camera lenses suffer from perceptible geometric distortion. To reduce this effect, we use the Brown-Conrady decentering distortion model catering for both radial and tangential distortions [35].

3.5 FEATURES EXTRACTION

At this point, we have obtained the rectangular image of every parking place, the related confidence field and we would like to extract the rele-

(a)

← occluded
part →

↑
unaffected area

(b)   (c)   (d)   (e)
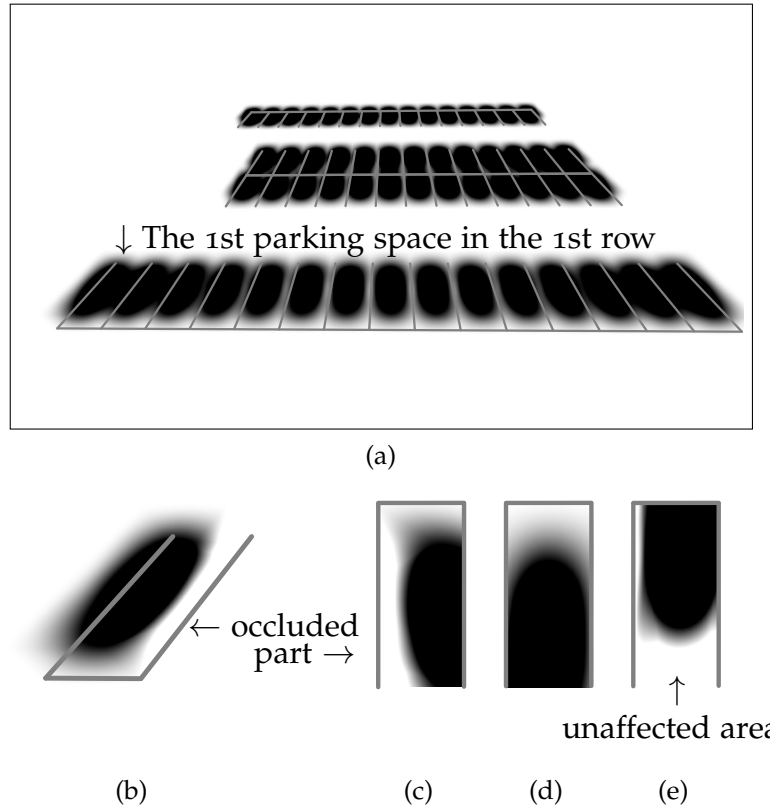
Figure 3.4: (a) The scalar field h representing the parking lot surface oc-
clusions. *Black level* refers to completely occluded parts of the
parking ground. (b) The scalar field c_1 represents the con-
fidence that the pixels may belong to a vehicle parked exclu-
sively on the first parking space. *Black level* refers to 100 % con-
fidence. (c) The projected version of the scalar field c_1. White
stripe on the side of parking space is due to the occluding ve-
hicle parked on the neighboring slot. (d), (e) Examples of two
other confidence maps for two different parking spaces. (e)
Bottom white area is completely unaffected by a well parked
car and should not be taken into account

vant features that would allow us to discriminate between the two possi-
ble states of parking spaces.

The ability to clearly discriminate the various kinds of objects is the
key point of the robust feature set. Local object appearance and shape
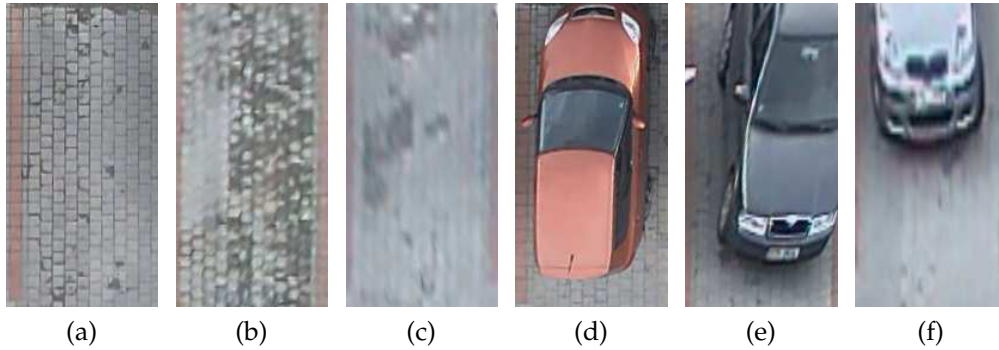
Figure 3.5: Normalized images of individual parking spaces. The images (a)-(c) contain parking spaces from three different parking rows viewed from different angles and distances. The same apply to the occupied parking spaces (d)-(f), but the problem is more pronounced due to the projection of car body parts outside the corresponding region. For the details about the parking lot arrangement, we refer the reader to Fig. 7.1

can often be characterized by the distribution of local intensity gradients or edge directions. Dalal and Triggs [45] showed in their experiments that the Histogram of Oriented Gradient (HOG) is one of the most successful edge and gradient-based descriptor and significantly outperforms existing feature sets for human or car detection. HOG is also very robust in the environments exhibiting large variations in appearances and illuminations. The method is based on evaluating the well normalized local histograms of image gradient orientations in a dense grid divided into small spatial regions called cells. The cells can be either rectangular or circular. Each cell is accumulating a weighted local 1D histogram of gradient directions over the pixels of the cell. Normalization improves the invariance to illumination and shadowing.

There also exist other well known and successful algorithms describing the local features like SIFT descriptors [100], scale and rotation invariant interest point detector and descriptor coined SURF [28], fast descriptor for real-time applications BRIEF [37], rectified Haar wavelets [112], Haar-like wavelets with AdaBoost [133, 134], low bit-rate descriptor CHoG [39], highly discriminative texture descriptor invariant to monotonic gray level changes LBP [138], SIFT-like descriptor that considers more spatial re-
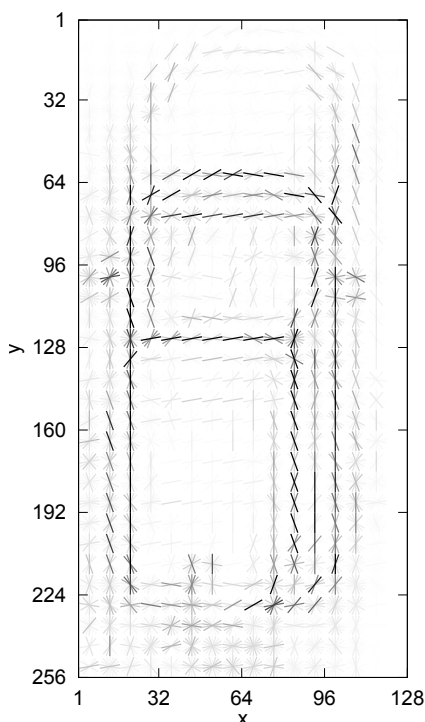
Figure 3.6: The results of HOG features extraction

gions for the histograms called GLOH. The comparison with many other descriptors of local regions can be found in [107].

Since we know the camera position we can obtain the unwarped (or normalized) image of every parking space and the related confidence field c_i. We extract the relevant features that would allow us to discriminate between the two possible states of parking spaces (see Fig. 3.6). We use 8×16 grids of 8×8 pixel cells each containing $\beta = 9$ orientation bins corresponding to evenly spread sectors of the half angle ignoring the direction. We will refer to the HOG of i-th rectangular patch of an image using the following vector notation $\text{hog}_i = (v_0, v_1, ..., v_{\beta-1}) \in \mathbb{N}^\beta$, where $\text{hog}(k) = v_k$ represents the number of votes for the k-th histogram channel or bin.

If we compare the image of an occupied parking space with an empty one, there is obvious difference in the distribution of prevailing edges. Simply put, the pertinence of the parking space to the given class may be devised from the total amount of cells, which can be regarded as the parts of parked car's edges. Moreover, based on the vector field **n** obtained by

projection of three-dimensional vector field of the car model iso-surface normals (e.g. for $h = 1$) onto the parking space image plane, we can roughly estimate the direction of such edges (i.e. the expected edge will be perpendicular to the local normal vector).

3.6 FEATURES PRIORITIZATION

(a)   (b)
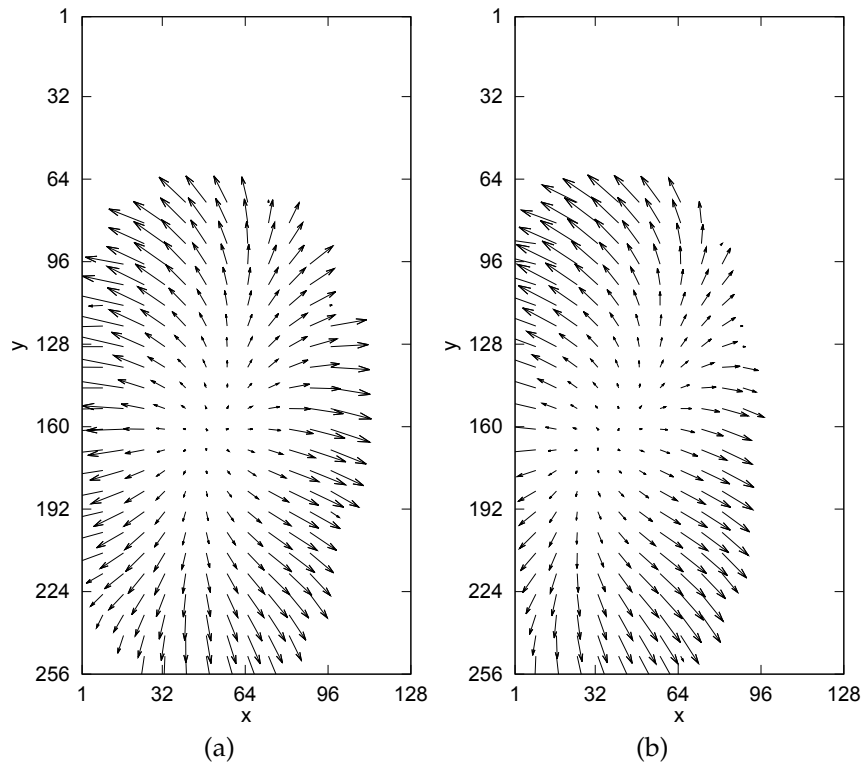
Figure 3.7: (a), (b) Planar vector fields represent the occurrence of a car hood edges in the normalized parking space image as viewed from the camera. Vector field is computed from the normals of the iso-surface of the probabilistic car model

Physically-inspired from the classical fluid dynamics, we may think of every HOG-cell as an idealized flowing fan-like mass object. Cells more conformal with the edge model will experience stronger drag force resulting in a higher velocity of these cells. As a result, these cells will be easily advected by the flow field from the origin position into the

31

detection zone. On the opposite side, cells with uniformly distributed bins will resist the flow. This will introduce the desired flexibility of our discrimination model with respect to the underlying car edge model. The motion of a Newtonian fluid with a constant density and temperature is governed by the Navier-Stokes equations (NSE) as follows

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \qquad (3.12)$$

where $\mathbf{u}$ represents the velocity vector field, p is the pressure field, ρ is the fluid density and ν is the kinematic viscosity of the fluid. The vector field $\mathbf{f}$ is an external force field and will be discussed in the following Section 3.6.1. In the case of incompressible fluids, the conservation of mass is then stated as the continuity equation $\nabla \cdot \mathbf{u} = 0$ meaning that the divergence of vector field $\mathbf{u}$ is zero. For the sake of brevity, the implementation details of solving the NSE are left uncovered. We adopt the approach thoroughly described in [126].

The NSE were successfully applied in many fields including image analysis and in our case, we interpret the resulting pressure and velocity fields as follows. The low pressure areas correspond to the sources of strong gradients caused by eventual car edges located at the positions predicted by the field of projected normals. The high pressure regions will represent the traps for moving particles. If the particle arrive in the detection area and has a strong dominant bin in the HOG, then we can suppose that the origin of the particle is placed somewhere close to a strong edge of a parked vehicle. The trajectory $\mathcal{P}$ in the conjunction with the actual bins configuration in particle's HOG should influence the speed of the moving particle. The new position $\mathbf{r}$ of the cell with the total mass m in the particular time step $t + 1$ is given by the formula

$$\mathbf{r}^{t+1} = \mathbf{F}^t \delta t^2 / m + 2\mathbf{r}^t - \mathbf{r}^{t-1}. \qquad (3.13)$$

The steady-state drag force $\mathbf{F}$ on the cell due to the fluid flow is derived from the standard quadratic drag equation for an object moving through a fluid and equals to

$$\mathbf{F}^t = \frac{1}{2}\rho\,\mathbf{u}\,||\mathbf{u}|| \max_{i \in \langle 0, b \rangle} \left\{ \left(1 - |\mathrm{rad2grad}\left(\mathrm{bin2rad}(i)\right) \cdot \hat{\mathbf{u}}|\right) C(\mathrm{hog}(i)) \right\},$$
$$\qquad (3.14)$$

no available data   no available data

(a)

in flow ⟶ ⟵   out flow   ⟶ ⟵ in flow

trap/detection area

(b)
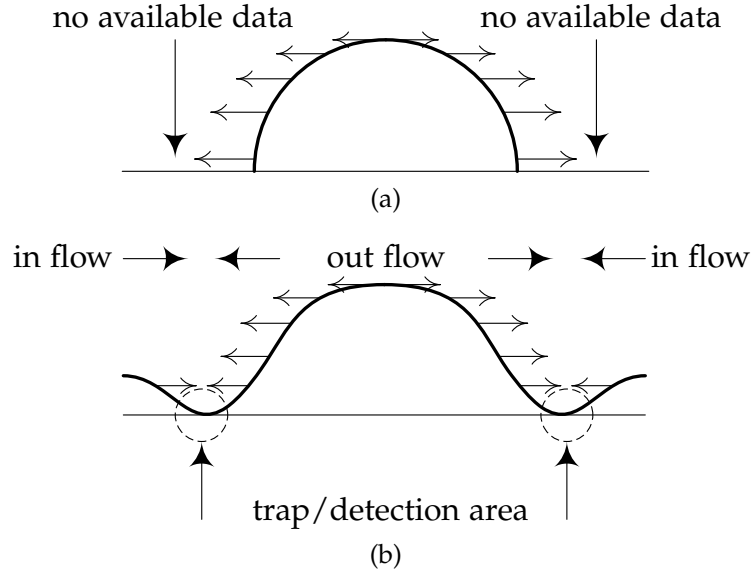
Figure 3.8: (a) Illustrative lateral section of the original force field as obtained from the car's hood normal field **n**, i.e. the only source of information about the expected shape (or surface) of a parked car. *Wide black* curves represent the speed of advection and the open arrows represent flow direction of features. (b) The resulting force field fulfilling the constraints imposed by the energy functional $\mathcal{E}$. Out flow is selectively expelling features from central area into the detection zone which is acting like a trap or a detector. Unwanted features by, in some sense, parasitic in flow are effectively attenuated by the confidence map

where the function rad2grad translates the angle in radians to unit direction vector and bin2rad converts i-th bin to radians. The constant b represents number of bins per orientation histogram. The hat over the **u** means that it is a unit vector and has magnitude equal to 1. The drag coefficient C is associated with the number of votes $v_i := \text{hog}(i)$ in the i-th histogram channel through simple polynomial function $C(v_i) = \alpha\, v_i^{\beta}$. For the rest of our experiments, the parameters were set as follows: $\alpha = 3$ and $\beta = 5$.
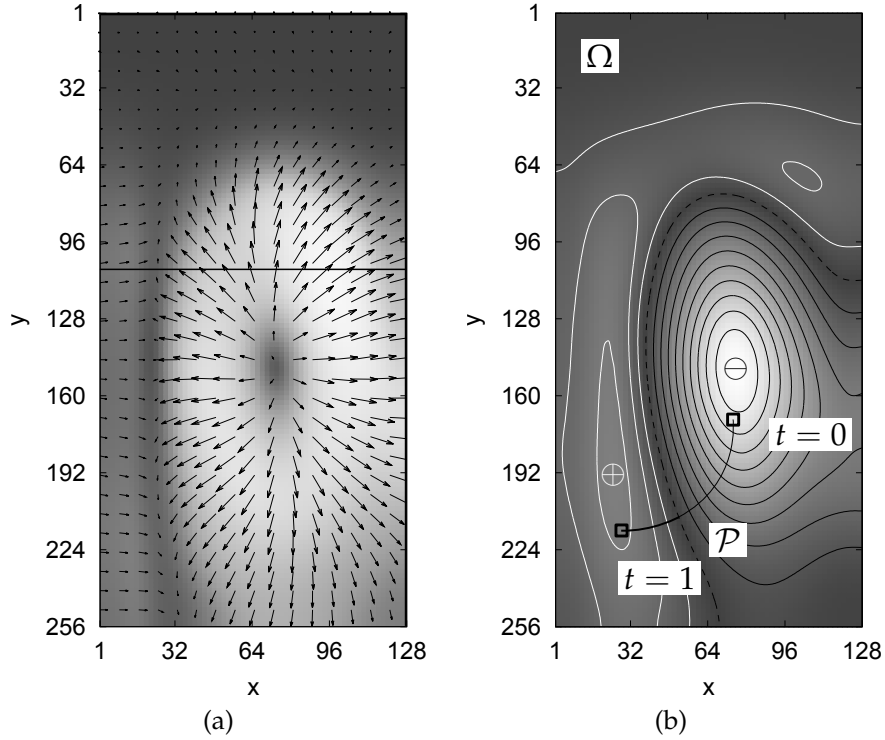
Figure 3.9: (a) An example of the resulting force field fulfilling the constraints imposed by the energy functional $\mathcal{E}$. *Horizontal black line* represents possible location of the lateral section from Fig. 3.8a. (b) The high-pressure field (*white iso lines*) marks the detection areas along the car boundaries. *Black box* represents a single advected HOG feature traveling across the simulation domain Ω and receiving votes

3.6.1 *Force Field Generation*

We expect that the external velocity field will start transferring the features from the regions of their abundance into the detection areas (see Fig. 3.8b). We can start with the gradient of the iso-surface of the h function which is subsequently projected on every parking place yielding a 2D vector field of normals **n**. The original normal vector field **n** is very close to fulfil the stated requirements on the field **f** which will initiate the motion of cells during the CFD steps. In order to assure that the force field fulfil the stated requirements even closer, we define the force field
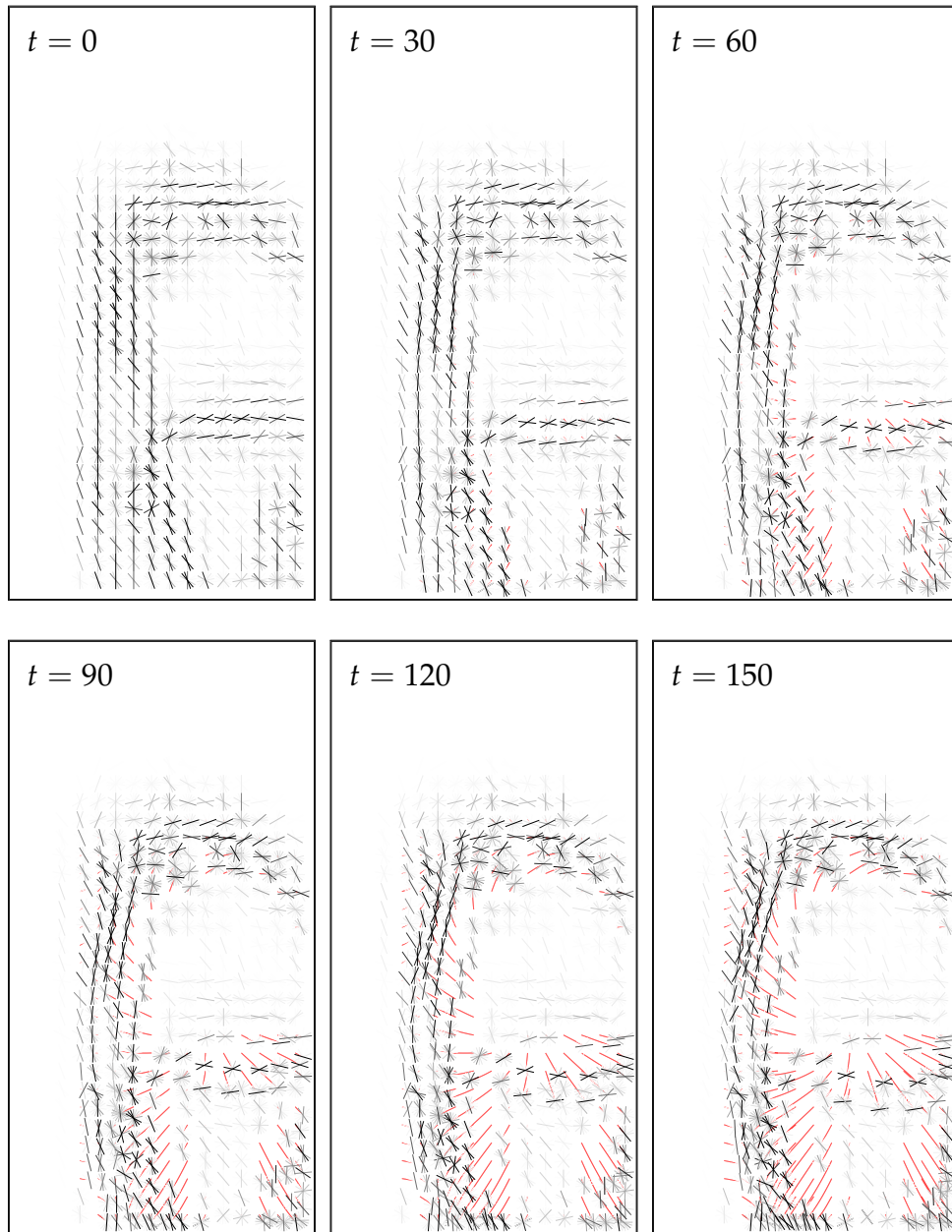
Figure 3.10: The series of six snapshots taken during features advection of the 30th parking space. *Solid lines* represent trajectories of individual cells. The majority of relevant features are accumulated in the detection area

to be the vector field $\mathbf{f}(x, y) = [u(x, y), v(x, y)]$ that minimizes the global energy functional

$$\mathcal{E} = \iint_\Omega \lambda_1 \left(||\nabla u||^2 + ||\nabla v||^2 \right) + \lambda_2 \, ||\mathbf{n}||^2 \, ||\mathbf{f} - \mathbf{n}||^2$$
$$+ \, ||\nabla c||^2 \, ||\mathbf{f} - \nabla c||^2 \, \mathrm{d}x \, \mathrm{d}y \, , \qquad (3.15)$$

where the first term in the functional follows a standard principle, that of making the result smooth when there is no data. The second term is the data attachment term, whose minimization tends to make the force field to be similar with the normal field especially in the areas where the normal field is large. The third confidence field c driven term enforces the presence of an in-flow from border areas and also partially helps to increase the pressure in the detection areas.

The system of corresponding Euler–Lagrange equations for unknown functions u and v of two real arguments x and y is

$$\frac{\partial}{\partial u}\mathcal{L} - \frac{\partial}{\partial x}\left(\frac{\partial}{\partial u_x}\mathcal{L}\right) - \frac{\partial}{\partial y}\left(\frac{\partial}{\partial u_y}\mathcal{L}\right) = 0$$
$$\frac{\partial}{\partial v}\mathcal{L} - \frac{\partial}{\partial x}\left(\frac{\partial}{\partial v_x}\mathcal{L}\right) - \frac{\partial}{\partial y}\left(\frac{\partial}{\partial v_y}\mathcal{L}\right) = 0, \qquad (3.16)$$

whose solutions are the functions for which a given functional $\mathcal{E}$ is stationary. Note that $u_x(x, y)$ denotes the partial derivative of u w.r.t. the variable x. Provided that

$$\frac{\partial}{\partial u}\mathcal{L} = \frac{\partial}{\partial u}\left(\lambda_2 \, ||\mathbf{n}||^2 \left((u - m)^2 + (v - n)^2\right)\right.$$
$$\left. + \, ||\nabla c||^2 \left((u - c_x)^2 + (v - c_y)^2\right)\right)$$
$$= 2\lambda_2 \, ||\mathbf{n}||^2 \, (u - m) + 2 \, ||\nabla c||^2 \, (u - c_x) \, ,$$
$$\frac{\partial}{\partial v}\mathcal{L} = 2\lambda_2 \, ||\mathbf{n}||^2 \, (v - n) + 2 \, ||\nabla c||^2 \, (v - c_y) \qquad (3.17)$$

and

36

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial u_x}\mathcal{L}\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial u_x}\left(\lambda_1\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right)\right)\right) = 2\lambda_1 u_{xx},$$

$$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial u_y}\mathcal{L}\right) = 2\lambda_1 u_{yy}, \quad \frac{\partial}{\partial x}\left(\frac{\partial}{\partial v_x}\mathcal{L}\right) = 2\lambda_1 v_{xx}, \quad \frac{\partial}{\partial y}\left(\frac{\partial}{\partial v_y}\mathcal{L}\right) = 2\lambda_1 v_{yy}.$$

$$(3.18)$$

Finaly, after applying the standard methods of the variation calculus and plugging Eqs. (3.17) and (3.18) into the system of Eqs. (3.16) we obtain two Euler-Lagrange equations

$$\lambda_2\left|\left|\mathbf{n}\right|\right|^2(u-m) + \left|\left|\nabla c\right|\right|^2(u-c_x) - \lambda_1\Delta u = 0,$$
$$\lambda_2\left|\left|\mathbf{n}\right|\right|^2(v-n) + \left|\left|\nabla c\right|\right|^2(v-c_y) - \lambda_1\Delta v = 0,$$

$$(3.19)$$

where Δ is the Laplace operator. Both Eqs. (3.19) can be solved iteratively by treating u and v as functions of time t according the time-marching scheme

$$\frac{\mathrm{d}}{\mathrm{d}t}u(x,y,t) = \lambda_2\left|\left|\mathbf{n}(x,y)\right|\right|^2(u(x,y,t) - m(x,y))$$
$$+ \left|\left|\nabla c(x,y)\right|\right|^2(u(x,y,t) - c_x(x,y)) - \lambda_1\Delta u(x,y,t),$$
$$\frac{\mathrm{d}}{\mathrm{d}t}v(x,y,t) = \lambda_2\left|\left|\mathbf{n}(x,y)\right|\right|^2(v(x,y,t) - n(x,y))$$
$$+ \left|\left|\nabla c(x,y)\right|\right|^2(v(x,y,t) - c_y(x,y)) - \lambda_1\Delta v(x,y,t).$$

$$(3.20)$$

These equations are decoupled, and therefore can be solved as separate scalar partial differential equations in u and v, provided that the partial derivatives with respect to time t on the left side of the Eqs. (3.20) are approximated by the first-order accurate forward difference formulas yielding

$$u^{t+1}(x,y) = u^t(x,y) + \delta t\left(\lambda_2\left|\left|\mathbf{n}(x,y)\right|\right|^2\left(u^t(x,y)\right.\right.$$
$$\left. - m(x,y)\right) + \left|\left|\nabla c(x,y)\right|\right|^2\left(u^t(x,y) - c_x(x,y)\right) - \lambda_1\Delta u^t(x,y)\Big),$$
$$v^{t+1}(x,y) = v^t(x,y) + \delta t\left(\lambda_2\left|\left|\mathbf{n}(x,y)\right|\right|^2\left(v^t(x,y)\right.\right.$$
$$\left. - n(x,y)\right) + \left|\left|\nabla c(x,y)\right|\right|^2\left(v^t(x,y) - c_y(x,y)\right) - \lambda_1\Delta v^t(x,y)\Big).$$

$$(3.21)$$

The iteration begins by setting $u^0(x,y) = m(x,y)$ and $v^0(x,y) = n(x,y)$. To ensure the convergence of the above described iterative process, we restrict the time step δt with the Courant-Friedrichs-Lewy (CFL) condition. An example of the resulting force field is shown in the Fig. 3.9a.

3.6.2 *Weight Assignment Procedure*

In this section, we will describe how to assign certain weight w_i to the individual cell (i.e. the rate of belonging to the car edge). As stated above, we track the position $\mathbf{r}$ of every cell as it moves across the simulation domain Ω represented by the normalized image of a parking space. In accordance with our cells advection model, the most relevant cells travel across high-pressure areas and should gain the most votes (or weight). This can be expressed by the following path integral

$$w_i = \int_{\mathcal{P}_i} \kappa(p(s)) \, \mathrm{d}s = \int_0^1 \kappa(p(\mathbf{r}(t))) \, ||\mathbf{r}_t(t)|| \, \mathrm{d}t \,, \qquad (3.22)$$

where $\mathcal{P}_i$ is the trajectory taken by the i-th cell due to the influence of the flow field $\mathbf{u}$ (see Fig. 3.9b). The factor $||\mathbf{r}_t(t)||$ represents the speed of traversal of the trajectory as the parameter t runs between two endpoints $t = 0$ and $t = 1$. The real-valued function κ converts the pressure into the weight value. We suggest to define this function as $\kappa(x) = \sqrt{\max(x,0)}$. This definition reflects our exclusive interest in the areas with positive pressure and also reduces the influence of the pressure magnitude on the resulting weight.

Taking the influence of the parking space distances into account.

The influence of distance of the parking space on the HOG features magnitude must be compensated to retain the possibility to discriminate the parking space state with only a single threshold. The actual threshold value depends on the mean of two exponential regression models approximating the scatter plot of both categories (see Fig. 3.11).

3.7 CRF TOOLKIT FOR 3-CLASS CELL LABELING

In our approach, parking space status inference can be considered as a labeling problem that involves assigning HOG cells (or sites) $\mathcal{S}$ a set of three
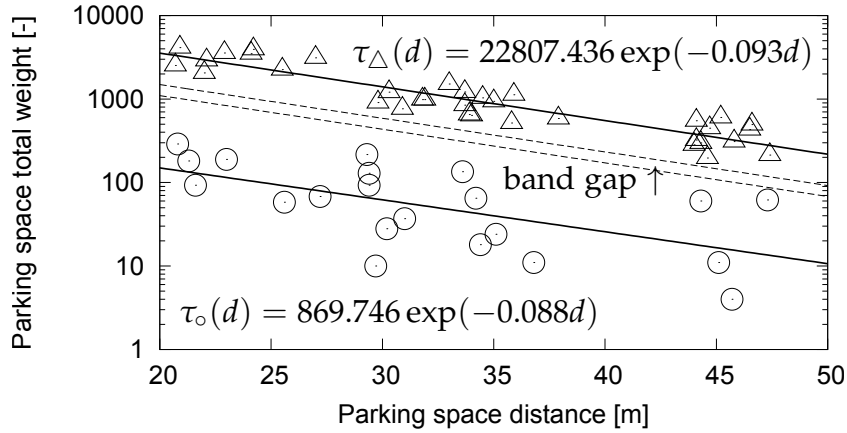
Figure 3.11: The plot with a logarithmic scale specified for the vertical axis shows the dependence of the parking space total weight $\sum_{\Omega} w_i$ on the parking space distance d from the camera. Two *solid lines* represent exponential regression models of measured (only the first frame of the first test sequence was used) total weights of vacant (*circles*) and occupied (*triangles*) parking spaces. The *dashed lines* denote the band gap where the state cannot be reliably assigned to either label

labels $\mathcal{L} = \{0 \equiv \text{no edge}, 1 \equiv \text{unknown}, 2 \equiv \text{car edge}\}$ with a subsequent decision Λ on the parking space state from the set $\Sigma = \{0 \equiv \text{vacant}, 1 \equiv \text{occupied}\}$. In other words, we seek for an optimal mapping $L : \mathcal{S} \to \mathcal{L}$ and a function $\Lambda : L \to \Sigma$. The mapping L is represented as a random field with the nodes aligned to the cells generated over the normalized image of a parking space. The random field $L = \{y_i : i \in \mathcal{S}\}$, where each random variable y_i takes on a value from the set of labels $\mathcal{L}$, is realized as a Conditional Random Field (CRF). The CRF is an undirected graphical model originaly developed for labeling sequential data [94]. Later, Kumar and Herbert [93] introduced the generalized discriminative framework for 2D images which allows the modeling of different types of interactions in labels and data.

Here, we want to find a configuration of L which maximizes the a posteriory (MAP) estimate of the underlying field given the observed data. This problem can be reformulated as a minimization of an energy function

$$E(\mathbf{y}; \mathbf{x}, \Theta) = \sum_{i \in \mathcal{S}} \psi_i^U(y_i; \mathbf{w}, \mathbf{c}, \Theta) + \sum_{(i,j) \in \mathcal{N}} \psi_{i,j}^P(y_i, y_j; \mathbf{h}), \qquad (3.23)$$

where the feature vector $\mathbf{x}$ contains acquired weights $\mathbf{w}$, representative colors of individual cells $\mathbf{c}$, and the HOG $\mathbf{h}$. Minimizing the energy function is known to be NP-hard problem. There exist the approximate solutions such as graph cut (st-MINCUT), generalized belief propagation (GBP), and the tree re-weighted message passing (TRW).

Maximum a posteriori probability estimate.

The most probable labeling $\mathcal{L}$ can be found by means of maximum a posteriori (MAP) inference or energy minimization

$$\mathbf{y}^\star = \arg \max_{\mathbf{y}} \mathbf{P}(\mathbf{y}|\mathbf{x}) = \arg \min_{\mathbf{y}} E(\mathbf{y}; \mathbf{x}, \Theta). \qquad (3.24)$$

All our experiments are performed with the loopy belief propagation (LBP). Very briefly described, new message is computed for every possible label in the following way

$$m_{p \to q}^{t+1}(y_q) = \min_{y_p \in \mathcal{L}} \left[\psi_p^U(y_p; \mathbf{x}) + \psi_{p,q}^P(y_p, y_q; \mathbf{x}) + \sum_{s \in \mathcal{N}_4(p) \setminus q} m_{s \to p}^t(y_p) \right]. \qquad (3.25)$$

After T iterations, when the stationary state is reached, the messages m do not change anymore, a belief vector b is evaluated for each node

$$b_q(y_q) = \psi_q^U(y_q; \mathbf{x}) + \sum_{p \in \mathcal{N}_4(q)} m_{p \to q}^T(y_q). \qquad (3.26)$$

The final step of Eq. (3.24) estimation is carried out by selection of pixel label minimizing the computed belief individually at each node

$$y_q^\star = \arg \min_{y_q \in \mathcal{L}} b_q(y_q). \qquad (3.27)$$

The time complexity of the above mentioned algorithm equals to $O(nk^2T)$, where n is the number of nodes, k is the number of possible labels, i.e. $k = |\mathcal{L}|$ and T is the number of iterations. More detailed description of the inference method is beyond the scope of this work and here we follow with the main ideas of the related potential function design.

40
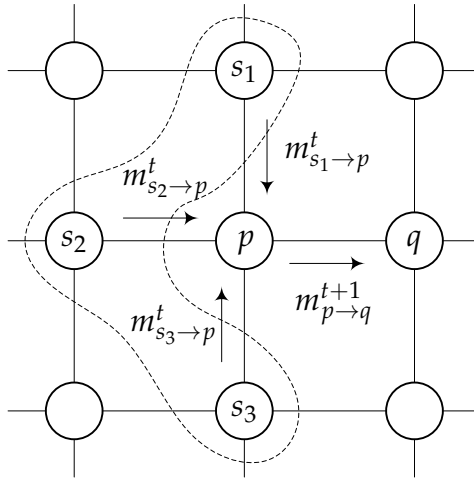
Figure 3.12: The max-product BP algorithm works by passing the messages around the graph defined by the 4-connected image grid. The method works in an iterative fashion, with the messages from all nodes being passed in parallel [59]. To update the message passing from p to q, we consider all messages flowing into p, except for the message from q

Unary potential.

We could simply assume that the parked car will introduce a lot of edges all around the normalized image of a given parking space. Alas, this assumption may be violated by a strong ground pattern. Therefore, to assign the site labeling, we rely on the weights w which should be more resistant to this phenomenon than the original image gradient magnitude. To verify the edge hypotheses, we also take into account the color $\mathbf{c}_i$ representing each cell. This yields the unary potential

$$\psi_i^U(y_i; \mathbf{w}, \mathbf{c}, \Theta) = \underbrace{-\log P_s(y_i|w_i, \Theta)}_{\text{CFD shape prior}} - \lambda_U \underbrace{\log \mathcal{L}_{y_i}(P_c(k_{\text{best}}|\mathbf{c}_i, \Theta))}_{\text{GMM color model}}, \quad (3.28)$$
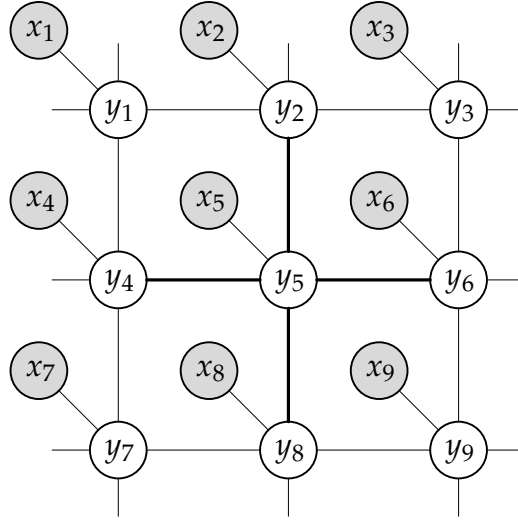
Figure 3.13: Conditional random field in the form of regular grid structure. The 4-connected neighbourhood of the 5th site is emphasized by *thick line*

that consists of two compounds. The first one represents the shape prior

$$P_s(y_i|w_i, \Theta) = \begin{cases} 1 - P(y_i = \text{ce}|w_i + \tau_{\text{ce}} - \tau_{\text{ne}}, \Theta) & , \ y_i = \text{ne} \\ 1 - (P_s(y_i = \text{ne}|w_i, \Theta) + P(y_i = \text{ce}|w_i, \Theta)) & , \ y_i = \text{u} \\ P(y_i = \text{ce}|w_i, \Theta) & , \ y_i = \text{ce} \end{cases} \quad .$$

$$(3.29)$$

where the posterior probability of labeling the i-th site (or cell) equals to a one-dimensional sigmoidal function as follows

$$P(y_i = \text{car edge}|w_i, \Theta) = \frac{1}{1 + \exp((\tau_{\text{ce}} - w_i)/s)} . \qquad (3.30)$$

The probability P_s relates weight values with the probabilities of particular labels (see Fig. 3.14 for further reference). The second compound P_c is the likelihood that the given cell color $\mathbf{c}_i$ match the most probable mixture component k_{best} of the actual GMM background model containing a mixture of $K = 2$ Gaussian densities (for shadowed and unshadowed regions). Besides the GMM, the set of global parameters Θ contains three experimentally evaluated constants: s is the steepness factor, and the threshold values $\tau_{\text{ne}}, \tau_{\text{ce}}$ are the inflection point of the sigmoid function for the no
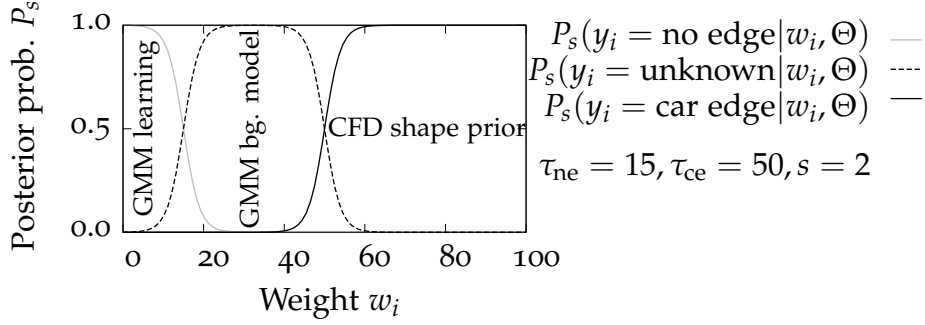
Figure 3.14: Posteriors used for defining the shape priors in the unary potential. The Π-shaped function (*dashed line*) represents the range of unreliable weight values originating from the CFD shape prior

edge label and car edge label. Translation functions $\mathcal{L}_{y_i}(x)$ return x, 0, and $1 - x$, respectively.

Pairwise potential.

In our approach, to define the pairwise relationship between two neighbouring cells, we need some criterion that will compare histograms of two given cells. Our dissimilarity measure is defined as follows

$$\phi(i, j) = \frac{4}{\beta^2} \sum_{k=0}^{\beta-1} \sum_{l=0}^{\beta-1} \frac{\text{hog}_i(k) \, \text{hog}_j(l)}{\left(\text{hog}_i(k) + \text{hog}_j(l) \right)^2} \frac{\rho(k, l)}{\rho_{\max}}, \qquad (3.31)$$

where the metric ρ returns the distance between two bins k and l of a histogram and is given by the formula

$$\rho(k, l) = \beta \left| u - \lfloor u \rfloor - \left\lfloor u - \lfloor u \rfloor + \frac{1}{2} \right\rfloor \right|, \qquad (3.32)$$

where $u = |k - l| / \beta$. The constant $\rho_{\max} = \lfloor \beta/2 \rfloor$ in Eq. (3.31) represents the maximum possible distance among all bins. Finally, the pairwise potential ψ^P is given by the multiplication of the dissimilarity measure and the delta function

$$\psi_{i,j}^P(y_i, y_j; \mathbf{h}) = \lambda_P \, \phi(i, j) \, \delta_{y_i, y_j}. \qquad (3.33)$$

We expect that such definition will favour the same labeling in the case of similar distributions of votes in the histograms of the cells with large weights. We use the loopy belief propagation (max-sum algorithm) to approximate the MAP inference and the state of the parking space is simply devised from the sum of the individual car edge labels of the final labeling (see Fig. 3.15).

3.8 EVALUATION AND CONCLUSION

In this section, we will evaluate the performance of two variants of the presented method; with and without contextual constraints imposed by the CRF. We observed 56 parking spaces in 4 rows with the network HD camera directly attached to the dedicated video server via Wi-Fi. The structure of the parking lot was measured and stored in an XML file (see Fig. 7.1). The camera was calibrated to match the parking lot geometry and also the geometric distortion (displacement of the pixel locations) caused by wide angle lens optics was suppressed. The evaluation is based on the series of seven video sequences (i.e. approx. 14 hours) captured during the most relevant parts of a day when a lot of cars are arriving or leaving monitored parking lot (see Fig. 7.2). The results of the variant without the background color model are summarized in the Table 3.1. Individual test images contain more than 10000 occupied parking spaces and more than 6000 vacant parking spaces. In addition to the table, false positive rate (FPR) and false negative rate (FNR) are no worse than 0.047 and 0.006, respectively. In comparison, other authors report the final false acceptance rate (FAR) 0.032 and false rejection rate (FRR) 0.020 [71]. Other method based on a SVM classifier achieves false detection rate (FDR) 0.048 and FRR 0.071 [48]. Furthermore, the results point out that the yaw angle φ compensation procedure (i.e. preserving the yaw of the cell constant as it moves along the path $\mathcal{P}$) is not the critical part of our algorithm. Columns In and Out represent the number of incoming, resp. outgoing vehicles. The second Table 3.2 describes experiments with another series of video sequences captured under different lighting conditions. In this case, the individual test images contain more than 14000 occupied parking spaces and more than 8000 vacant parking spaces in total. Our performance evaluation includes confusion matrix, F1 score, and Matthews correlation coefficient.

44
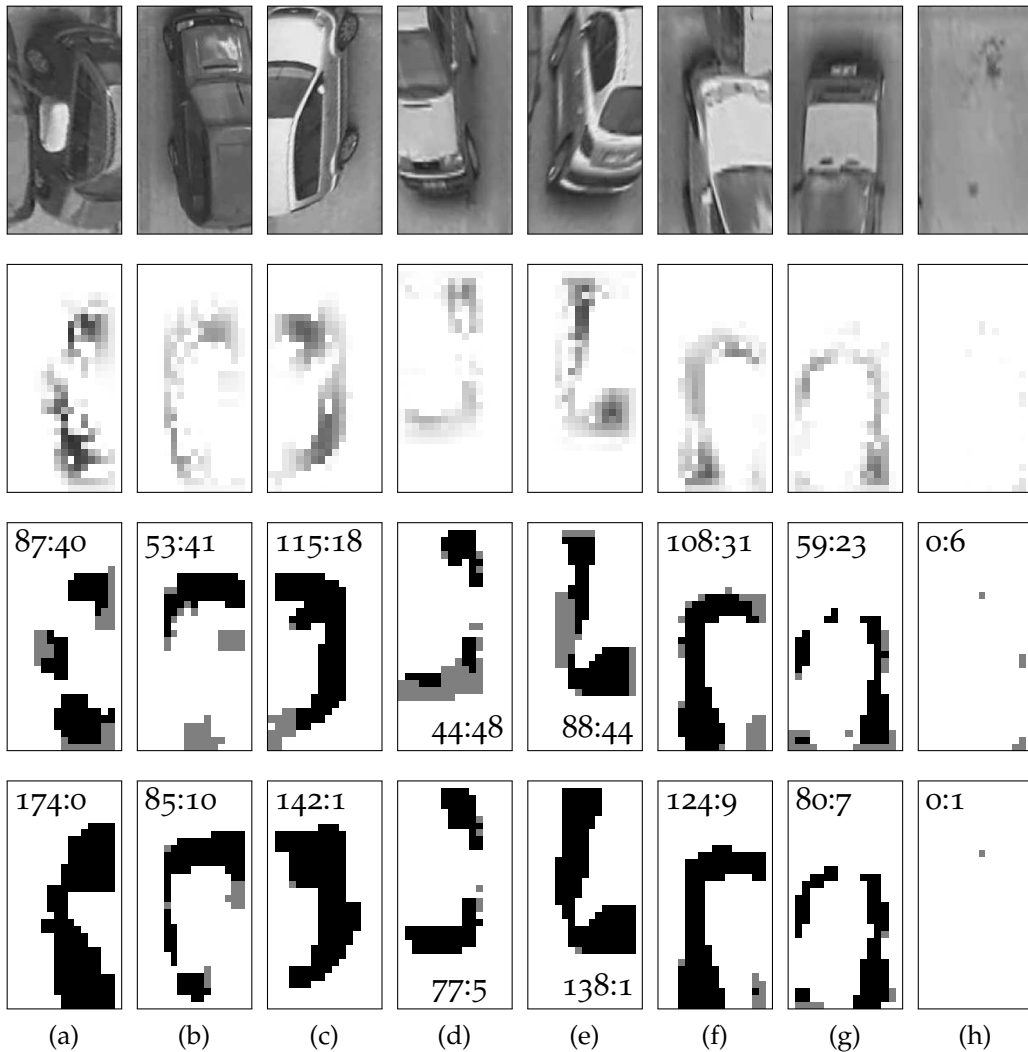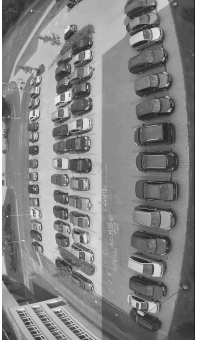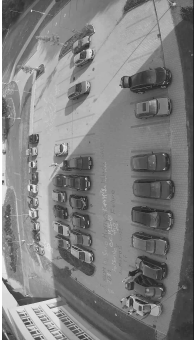
Figure 3.15: (first row) Examples of normalized images, (second row) corresponding weight maps of selected edges, and (third row) labeling obtained via the CRF without the background term. (h) Empty parking spaces containing only a few cells labeled as unknown (*gray*, the second number) provide the reliable sources of color patches for generating and updating the background GMM. (fourth row) When enabled, the background color model strengthens the difference between vacant and occupied parking spaces in terms of number of cells labeled as car edges (*black*, the first number). The parameters are: $\lambda_U = 0.45$, $\lambda_P = 50$
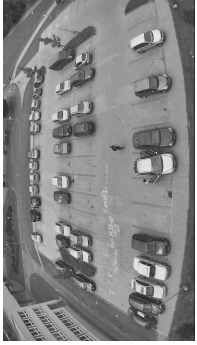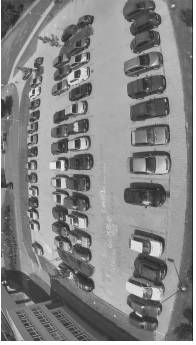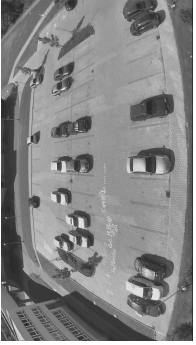
Table 3.1: Results of our algorithm without CRF as compared against the ground truth data (sequences #1-#3)

Original image	Row	TP	FN	FP	TN	In	Out	Acc.	Recall	Prec.	Spec.	F1
	1	600	0	18	1179	1	8	0.990	1.000	0.971	0.985	0.985
	2	494	0	1	1065	0	5	0.999	1.000	0.998	0.999	0.999
	3	609	0	3	948	2	8	0.998	1.000	0.995	0.997	0.998
	4	1230	12	14	539	7	7	0.986	0.990	0.989	0.975	0.990
	All	2933	12	36	3731	10	28	0.993	0.996	0.988	0.990	0.992
φ compensation		2937	9	37	3729			0.993	0.997	0.988	0.990	0.992
	1	915	0	65	817	18	3	**0.964**	1.000	**0.934**	0.926	**0.966**
	2	1204	29	1	316	15	2	0.981	**0.976**	0.999	0.997	0.988
	3	1268	0	0	292	12	2	1.000	1.000	1.000	1.000	1.000
	4	1515	0	18	267	15	3	0.990	1.000	0.988	0.937	0.994
	All	4902	29	84	1692	60	10	0.983	0.994	0.983	0.953	0.989
	1	666	0	20	205	12	0	0.978	1.000	0.971	**0.911**	0.985
	2	610	1	2	166	9	0	0.996	0.998	0.997	0.988	0.998
	3	525	1	4	238	10	1	0.993	0.998	0.992	0.983	0.995
	4	746	0	4	147	9	1	0.996	1.000	0.995	0.974	0.997
	All	2547	2	30	756	40	2	0.990	0.999	0.988	0.962	0.994

Table 3.2: Results of our algorithm with CRF as compared against the ground truth data. This sequences (#4-#7) contains inferior lighting conditions compared to sequences #1-#3

Original image	Row	TP	FN	FP	TN	Acc.	Recall	Prec.	Spec.	F1	MCC
	1	584	15	19	1181	0.981	**0.975**	0.968	0.984	0.972	0.958
	2	486	7	1	1065	0.995	0.986	0.998	0.992	0.992	0.988
	3	606	2	2	949	0.997	0.997	0.997	0.998	0.997	0.995
	4	1235	11	14	540	0.986	0.991	0.989	0.975	0.990	0.967
	All	2911	35	36	3735	0.989	**0.988**	0.988	0.990	0.988	0.979
	1	1667	0	13	120	0.993	1.000	0.992	**0.902**	0.996	0.946
	2	1416	0	0	144	1.000	1.000	1.000	1.000	1.000	1.000
	3	1363	0	12	184	0.992	1.000	0.991	0.939	0.996	0.965
	4	1695	0	3	102	0.998	1.000	0.998	0.971	0.999	0.985
	All	6141	0	28	550	0.996	1.000	0.995	0.952	0.998	0.973
	1	1248	0	57	495	0.968	1.000	0.956	0.897	0.978	0.926
	2	579	0	1	979	0.999	1.000	0.998	0.999	0.999	0.999
	3	975	0	3	582	0.998	1.000	0.997	0.995	0.998	0.996
	4	927	1	84	781	**0.953**	0.999	0.917	0.903	0.956	**0.909**
	All	3729	1	145	2837	**0.978**	1.000	**0.963**	**0.951**	**0.981**	**0.957**
	1	353	0	33	364	0.956	1.000	**0.915**	0.917	**0.955**	0.916
	2	423	1	0	226	0.998	0.998	1.000	1.000	0.999	0.997
	3	392	0	9	247	0.986	1.000	0.978	0.965	0.989	0.971
	4	627	4	1	118	0.993	0.994	0.998	0.992	0.996	0.975
	All	1795	5	43	955	0.983	0.997	0.977	0.957	0.987	0.963

Table 3.3: Comparison of our algorithms against the selected methods from the state of the art Chapter 2. Ranking of the methods is based on the average accuracy

Method	Sequence	FAR	FRR	Acc.	Acc. Ranks
Fabian [4]	1	0.0013	0.0055	0.993	
	2	0.0043	0.0125	0.983	
	3	0.0006	0.0090	0.990	
	Avg	0.0021	0.0090	0.989	#2
Fabian [3]	4	0.0052	0.0054	0.989	
	5	0.0000	0.0042	0.996	
	6	0.0001	0.0216	0.978	
	7	0.0018	0.0154	0.983	
	Avg	0.0018	0.0116	0.987	#3
Huang [70]	8	0.0004	0.0081	0.999	
	9	0.0024	0.0324	0.996	
	10	0.0040	0.0437	0.994	
	Avg	0.0023	0.0281	0.9963	#1
Huang [71]	8	0.0004	0.1690	0.983	
	9	0.0002	0.2626	0.985	
	10	0.0042	0.1019	0.992	
	Avg	0.0016	0.1778	0.9865	#4
Wu [141]	8	0.0111	0.7115	0.919	
	9	0.0016	0.7837	0.958	
	10	0.0018	0.7012	0.974	
	Avg	0.0048	0.7321	0.9503	#5
Dan [48]	8	0.0307	0.5748	0.915	
	9	0.0101	0.7061	0.954	
	10	0.0073	0.6524	0.970	
	Avg	0.0160	0.6444	0.9464	#6

HOG FEATURES STABILIZATION

As we stated above, visual surveillance systems include a wide range of related areas ranging from motion detection, moving object classification and tracking to activity understanding. Each of the above-mentioned applications relies greatly on proper motion segmentation method. Many background subtraction algorithms have been proposed. Simple yet robust frame differencing, statistically based Mixture of Gaussians (MoG) which also bear the name of Gaussian mixture model (GMM) in the literature, sophisticated methods based on wavelets or the optical flow computed by the finite element method. In this chapter, we focus on novel modification of well known MoG. The intrinsic motivation stems from the inability of regular MoG implementation to handle many camera related phenomena. Here presented method exploits Histograms of Oriented Gradients to significantly reduce the influence of camera jitter, automatic iris adjustment or exposure control causing severe degradation of foreground mask. The robustness of introduced method is shown on series of video sequences exhibiting mentioned phenomena.

CITATION The method presented here was published in [2] and is cited in the large survey of statistical background modeling methods [33] and in conference papers [148, 27]. This approach is qualified as HOG feature improvement of the classical MoG.

4.1 MOTIVATION AND RATIONALE

During the last two decades, we can see a huge development in area of video surveillance. Many successful approaches and algorithms had appear but in spite of the fact that this area of image processing is extremely various in terms of working environments and objects of interest, there is still a lack of system capable of working in real-world conditions. In this section, we focus on a common approach to identify the moving objects

– a background subtraction and we introduce a novel method for robust background modeling.

The modeling of background image plays a very important role in the general process of background subtraction [42]. The resulting foreground mask is processed afterward by various high–level algorithms realizing desired functionality of surveillance or tracking systems used in various places and environments. The fruitfulness of the system depends greatly on quality of foreground mask which discriminate moving foreground from static background. Many background subtraction methods work on per-pixel basis, e.g. Wren *et al.* [139] treats every pixel as an independent Gaussian random variable, Stauffer and Grimson [127] proposed the mixture of 3 to 5 Gaussians to represent multi-modal pixel distributions, Elgammal *et al.* [54] introduced the use of kernel density estimation to exploit statistical correlation among neighbor pixels. Mentioned methods suffer from lower foreground mask quality during sudden camera jitter or auto-exposure control intervention. Preprocessing video sequence before segmentation is inefficient while stabilization algorithms for camera motion compensation are unable to restrain sub-pixel motion sufficiently in most cases. The camera vibration may also suffer from rotation and scaling as nonlinear motions that may embarrass the matching process [113].

There also exist methods reflecting the naturalness of pixel adjacency. Im *et al.* in [92] propose wavelet based moving object segmentation using high frequency coefficients in wavelet subbands. Antić *et al.* [26] propose a novel wavelet based method for robust feature extraction and tracking. They also claim, that extremely harsh conditions can occur and violate the premises of the statistical regularity and predictability of background pixels.

The goal of this chapter is to eliminate the unfavourable influence of camera related phenomena like sudden changes of overall image brightness caused by the auto-exposure control and to overcome camera jitter from unstable support. The rest of this chapter is organized as follows. The algorithm is described in Section 4.2. Evaluation and experimental results are shown in Section 4.3. Conclusions are given in Section 4.4.

Figure 4.1: An overview of proposed background subtraction method with stabilization

4.2 ALGORITHM OF STABILIZATION

In previous section, we mentioned two main problems associated with camera jitter and automatic exposure control. To reduce the unwanted influence of such events on quality of foreground mask, we replace the brightness value acting as a random variable in regular MoG algorithm by a local image gradient. The motivation behind this step is quite easy. Image gradient, and especially its orientation, is invariant against changes in overall image brightness. Of course, this applies only to a certain extent given by limited range of pixel values. In practice, this assumption holds quite well and we are able to reduce the aftereffects of automatic control over exposure that way. But this still will not help us to achieve the independence of foreground mask on camera jitter. We also need to handle small image movements. To do so, we evaluate the gradient not for every single position in the image, but for small squared area, e.g. 8×8 pixels called cell. Further, we introduce the on-line spatial rearrangement of cells to minimize the variance of dominant gradient for every cell.

We shall discuss the first step of our approach in more details now. We need to estimate the locally dominant gradients computed on a dense grid of uniformly spaced cells. Histograms of Oriented Gradients (HOG) [44] suits our needs perfectly, i.e. we use $a \times b$ grids of $c \times c$ pixel cells each containing β orientation bins corresponding to certain sector to the full angle. We will refer to the HOG of particular rectangular area i, j of image f using the following notation $\text{hog}_{i,j}(\nabla f) \rightarrow (v_0, v_1, ..., v_{\beta-1}) \in \mathbb{N}^\beta$, where v_k represents weighted votes for individual angle spans. The weight is proportional to the gradient magnitude. The very first step in HOG evaluation is to estimate the image intensity gradient ∇f in every

pixel of input image f. More specifically, the five point central difference formula is used to calculate the image gradient.

Now we need to evaluate HOG for every cell in the image. This is quite straight forward process briefly described in previous paragraph. For more details about computing HOG the reader is referred to [44]. As a result we obtain the most significant orientation bin for every cell in the image.

The computation of the mean HOG, denoted by $\text{mhog}_{i,j}$, follows. The purpose of mean histogram is to provide the reference values for minimizing the variance of gradients and especially resulting bins. We suppose that proper cell-shift vector Δx effectively eliminates camera jitter and other high frequency noise. If the whole image undergoes some small movement or is locally disturbed by some opto-physical process then obtained cell-shifts help us to stabilize the dominant bin values across the time. We also suppose that this will not harm the desired segmentation ability as the changes of the image gradient caused by transit of detected object will be much higher. This can be formalized by the following expression

$$\arg \min_{\Delta x}(\text{var} \underbrace{\left(\max\left(\text{hog}_{i,j}\left(\nabla f_t\left(x + \Delta x\right)\right)\right)\right)}_{\text{dominant gradient bin}}).\qquad(4.1)$$

The validity of both assumptions is discussed in Section 4.3. To estimate the most probable distribution of bins across the histogram we use on-line averaging as follows $\text{mhog}_{i,j}^t = (1 - \alpha)\,\text{mhog}_{i,j}^{t-1} + \alpha\,\text{hog}_{i,j}^t$, where α is the learning rate. Obtained histogram mhog represents the characteristic distribution of gradient orientations for a particular image area during a certain period of time. The similarity of both values mhog and hog indicates the presence of almost the same pattern in corresponding cell. We are looking for the best match following the spiral like trajectory until the similar bin is found or the border of search area is reached. In the later case we use the shift with smallest distance to the reference value.

Optimal cell-shift can be alternatively found by well-known Lucas-Kanade image alignment algorithm [102]. Just recall that in this case we are looking for alignment minimizing the difference between template image and an input image so that $\sum_x \left[f(W(x,p)) - f(x)\right]^2$, where W represents the

Table 4.1: Example of the distance matrix of metric ρ_1

ρ_1	0	1	2	3
0	0	1	1	1
1	1	0	1	1
2	1	1	0	1
3	1	1	1	0

Table 4.2: Example of the distance matrix of metric ρ_2 for particular 4-bin histogram

ρ_2	0	1	2	3
0	0	1	2	1
1	1	0	1	2
2	2	1	0	1
3	1	2	1	0

set of allowable warps parameterized by p. For the simplest case of pure translation motion it holds that $p = \Delta x$.

To measure the distance between two different bins, we need to introduce some reasonable metrics. We have started with very simple metrics in a discrete metric space defined as follows

$$\rho_1(x,y) = \left\{ \begin{array}{ll} 0 & x = y \\ 1 & x \neq y \end{array} \right. . \tag{4.2}$$

The obvious property of metrics (4.2) is it makes no difference between two different bins. Two different bins are equally good and don't matter if they are very similar or completely different. The positive effect of such simplified metric is that the evaluation is very fast. The apparent disadvantage is that in situation when a same bin doesn't exist (and it is true for most cases) we have to make an ad-hoc decision (e.g. we pick the first one). Due to this reason we have defined more complex metric as follows

$$\rho_2(x,y) = \left| uv - v \lfloor u \rfloor - v \left\lfloor u - \lfloor u \rfloor + \frac{1}{2} \right\rfloor \right|, \tag{4.3}$$

where $u = |x - y| / v$ and $v = \beta$. With metric (4.3) we are able to find the best solution of Expression 4.1 among all valid cell-shifts. As stated before, such solution minimizes the variance of winning bin for each cell and we effectively reduce the unwanted influence of camera jitter. Typical example of resulting cell-shift can be seen in Fig. 4.3.

Now we need to incorporate the dominant bin into the MoG scheme. The general MoG is very well described in [127, 114] and the related

Table 4.3: Essential values of presented method's parameters used during evaluation

a	b	c	r	α	β	metric	K	σ_{init}	ω_τ
80	60	8	5	0.001	16	ρ_2	9	500.0	0.3

maximum-likelihood parameter estimation problem can be found in [29], thus we will not repeat here the whole theory again and we restrict ourselves to only those parts that are different. Authors in [127] consider the values of particular pixels over time as a "pixel process". Similarly in our approach we can talk about an "edge process" which consist of time series of observations $\{X_0, X_1, ..., X_t\}$, where $X_i = \text{hog}^i$. The process is modelled by a mixture of K Gaussian densities with the set of parameters Θ_k, one for each state k

$$f_{\mathbf{X}|k}(X|k, \Theta_k) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k)}, \tag{4.4}$$

where $\Theta_k = \{\mu_k, \Sigma_k\}$ and Σ_k is the full covariance matrix (see Fig. 4.2). For computational reasons and the assumption that the red, green, and blue pixel values are independent and have the same variances, many authors assume the covariance matrix to be of the form $\Sigma_k = \sigma_k^2 \mathbf{I}$. This is not our case as we need to maintain the ability of general Gaussian distribution to represents the elongated data sets (as the orientation of bin is likely to be more stable than the gradient magnitude).

The rest of MoG implementation is same as described in [114]. Estimation of parameters μ_k and Σ_k strictly follows [114, 29]. The segmentation phase depends on proper estimation of value ω_τ which represents the classification threshold. If the probability of certain surface k is higher than ω_τ then the surface is regarded as a background.

4.3 PERFORMANCE EVALUATION

In this section, we need to define how to assess the quality of foreground detection. Proposed algorithm, as well as other background subtraction methods, produces a binary image in which we need to identify correctly identified pixels. To do so, we have generated a ground truth mask for
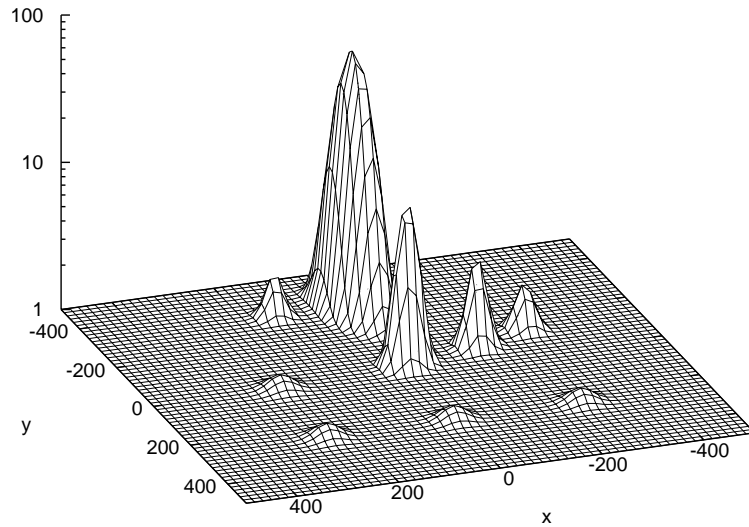
Figure 4.2: An example of mixture of 9 Gaussians. Snapshot was taken during the early stages of background model adaptation so the modes are still organized into the initial regular grid. It is also apparent that corresponding cell covers an image area without the presence of a significant edge. Also the range of image gradient is limited

three test sequences. Sequences were captured under different lighting conditions. Manual generation of ground-truth data in case of long sequences is too difficult and time-consuming, so we have decided to generate reference mask in the following way. Test sequences were captured under nearly ideal conditions, i.e. no automatic exposure adjustment nor exposure change nor camera jitter has occurred. In this case, we can use even simple background subtraction algorithm (e.g. frame differencing with carefully selected single frame acting as a reference background image) to obtain high quality foreground mask. During the testing phase we apply predefined filters on original video sequence to simulate desired camera related phenomena. Namely, we use translation to simulate camera jitter and we also perform both linear and nonlinear point operations to introduce the effect of automatic camera exposure control (for further details refer to the Fig. 4.4). In this way, our approach allows us to control the amount of spurious effects introduced into the test sequences. Other authors propose the use of semi-synthetic ground-truth sequences

(a)   (b)
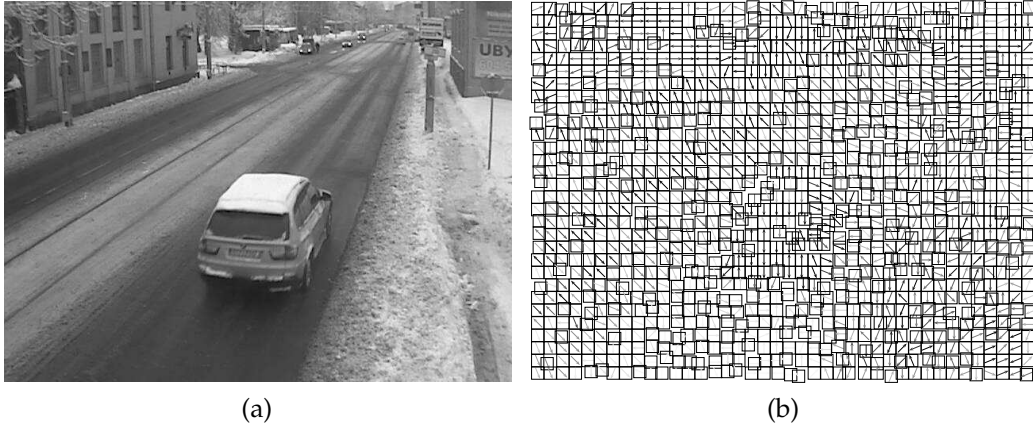
Figure 4.3: (a) Sample frame from the test sequence and (b) the corresponding histograms of oriented gradients; only the most significant bin for each cell is shown. Note the different shifts of cells minimizing the time-variance of dominant bins. Position of every cell obey the minimization criterion of Expression (4.1), hence it will reduce the influence of camera jitter as well as the high frequency noise

where previously segmented tracked objects are artificially inserted into real video sequences.

Table 4.4: Contingency table of four possible conclusions that can be drawn in a statistical hypothesis test

Real output	Ground Truth	
	Foreground	Background
Foreground	True Positive (TP)	False Positive (FP)
Background	False Negative (FN)	True Negative (TN)

We adopt two types of pixel-based metrics. The first one comes from [128]. The absolute error e_a is defined as follows

$$e_a = \frac{N_{FP} + N_{FN}}{a \cdot b}.$$ (4.5)
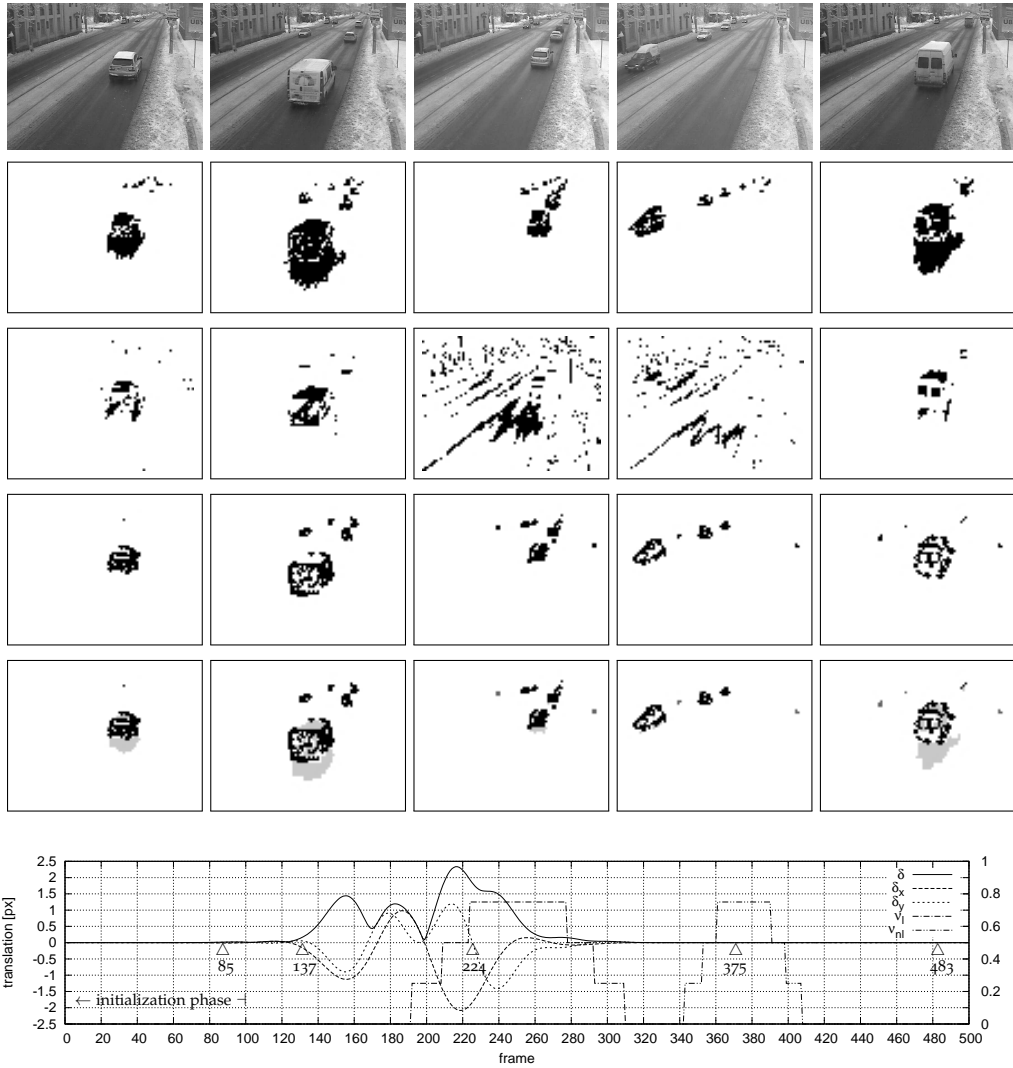
Figure 4.4: (Continued on the following page . . .)

Figure 4.4: Evaluation of proposed method. The first row represents five selected frames from the first video sequence captured by the stationary camera, the second row shows reference foreground mask, the third row contains MoG output, the fourth row presents proposed method's output, and the fifth row compares output of presented method with reference mask (black - TP, white - TN, light grey - FN, dark grey - FP). It is apparent, that the original MoG cannot handle the sudden changes of pixel values and the proposed method generates valid foreground mask (frames 224 and 375). Note that the reference segmentation also includes shadows, what is slightly inappropriate, since the proposed method correctly evaluates shadowed areas as a background. The plot at the bottom of the figure shows values of three coefficients; δ, resp. δ_x and δ_y - image translation simulating camera jitter, v_l - linear point operation (brightness change equals to $30 \cdot v_l$), v_{nl} - nonlinear point operation (gamma correction with $\gamma = 1/(1 + 0.5v_{nl})$). Positions of selected frames are also marked. Initialization phase takes 100 frames and the learning rate α is 100-times higher during this period

For the meaning of subscripts of number of pixels N please refer to the Table 4.4. We recall that a and b stand for foreground mask dimensions. The second metric e_g is defined subsequently

$$e_g = \frac{1}{|\Omega_f|} \sum_{x \in \Omega_f} \min_{y \in \Omega_f^1} ||x - y||_2, \tag{4.6}$$

where Ω_f represents the set of foreground pixel coordinates in actual foreground mask, Ω_f^1 stands for the set of foreground pixels in reference foreground mask. Simply put, metrics e_g returns the average minimum distance between all pixels of the current mask and the foreground pixels of the reference mask. Both metrics are calculated for each frame and overall values are obtained by averaging over the entire test sequence.

Another evaluation metrics can be derived from Table 4.4; true positive rate and false positive rate. A receiver operating characteristic (ROC), or simply ROC curve, enables evaluation of optimal parameter settings for

generic algorithm (a binary classifier), and the objective comparison of two or more algorithms [95]. In Fig. 4.5 we show that the TPR and FPR values are sensitive to the presence of noise in the foreground mask.
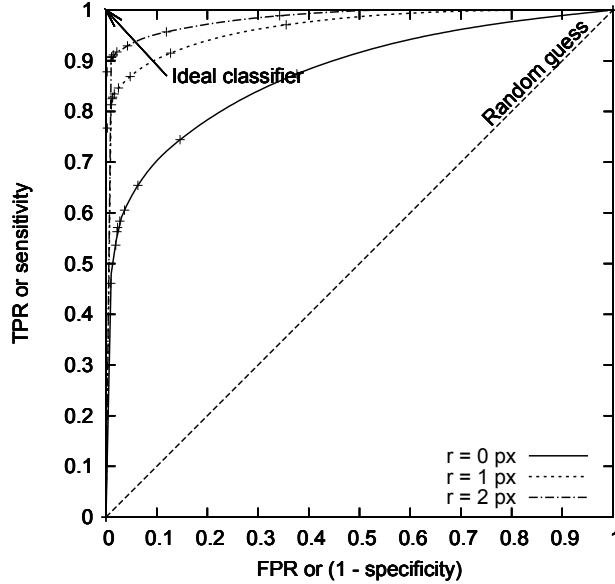
Figure 4.5: Receiver operating characteristic for three different search radii (0, 1 and 2 pixels). The plot shows that larger search radii lead to better results

The most important columns in Table 4.5 are FP and e_g. We can see that the number of pixels falsely marked as a foreground is significantly lower in the case of presented method and direct observation from Fig. 4.4 is consistent with the measured values. Higher values of FN can be explained by the presence of shadows in reference foreground mask.

4.4 CONCLUSION

In this chapter, we have proposed a new algorithm for background subtraction which is based on two well known methods: Mixture of Gaussians and Histograms of Oriented Gradients. Resulting method is significantly less vulnerable to the most common camera related phenomena like jitter and automatic exposure control. Higher level algorithms may benefit from the more reliable foreground mask and may track the moving objects even during the moments, when the ordinal MoG is failing.

Table 4.5: The results of original MoG and proposed method compared with the reference background mask

	Method	TP	TN	FN	FP	e_a	e_g
seq. 1	MoG	95.8	4462.5	146.5	95.2	5.0 %	1755.5
	MoG + HOG	82.1	4528.3	160.2	29.4	3.9 %	81.7
seq. 2	MoG	398.8	3637.2	665.8	98.2	15.9 %	832.4
	MoG + HOG	333.7	3657.0	730.9	78.3	16.9 %	199.0
seq. 3	MoG	44.3	4651.5	63.2	41.0	2.2 %	692.8
	MoG + HOG	55.3	4669.2	52.2	23.3	1.6 %	186.3

Tests performed on three video sequences with a total length of 1500 frames confirmed the assumptions made in the introductory section. Presented method is computationally more expensive then ordinal MoG. Focus should be placed on more effective implementation of this method and especially on evaluation of the gradient histograms and probability density functions with full covariance matrix. In the context of the parking lot surveillance, we have used this approach in two directions. Firstly, we have used this algorithm to stabilize the obtained features and secondly, we utilize devised metrics to compute the bin distance of two hog-cells.

DIRECT VOLUME RENDERING

In this section, we attempt to solve the so called standard volume rendering integral (VRI) from the Section 2.4. In the our context, we need to solve this integral to obtain probabilistic maps of occlusions (see Fig. 3.4b). But in general, the VRI is the merit of the direct volume rendering (DVR), the area of computer graphics dealing with visualization of scalar data. The most common approach is to approximate the integral by a Riemann sum with a fixed step size. Contrary to the traditional approaches, we try to devise its solution in an analytical form. Our volume rendering method employs trilinear interpolation of a discrete data set to reconstruct a cubic polynomial of densities along the ray segment passing through each cell of the volume. We obtain analytical expression of the densities across the ray-volume intersection segment. This form allow us to simply evaluate the accumulation of density along the ray path simply by taking the integral of polynomial. First order differential quantity like analytical gradient is provided as well. Results from this chapter are also applicable for computing vector fields (see Fig. 3.9a). Furthermore, we also define an emission term which allow us to carry out an closed form of the aforementioned integral for a simple case of Lambertian reflectance with a single directional light. As a result, we obtain an image of analytically ray casted structures inside the volume.

5.1 INTRODUCTION

Direct volume rendering (DVR) is a wide spread method for direct visualization, analysis and classification of the structures inside scalar volume data sets. Its applications range from medical imaging, oil and gas industry, mechanical engineering, fluid dynamics, astrophysics, chemistry to various special effects in the film industry and games. Volume data sets are often acquired by scanning the using magnetic resonance imaging (MRI), computational tomography (CT), positron emission tomography (PET), and also by physical simulations. DVR may be regarded as a ma-

ture technique since the publication of the pioneering works of Levoy [97], Drebin et al. [51] in late 1980s. In comparison to iso-surface rendering of 2-dimensional manifolds, the DVR deals with volume as a participating medium in which light can be absorbed, scattered, or emitted, as it passes through the volume of scalar values organized into structured on unstructured grid. Optical properties of various materials captured in the volume are classified by the so called transfer function.

Basically, there exist four main techniques of volume rendering: volume ray casting, splatting, hybrid strategies such as the shear warp algorithm, and texture-based rendering. Volume ray casting is considered to be the most versatile method providing very high quality images. In short, clipped ray passing the volume is sampled at regular or adaptive intervals. Interpolated data are plugged into the transfer functions and obtained color and opacity values are successively composed in front-to-back or back-to-front scheme.

5.2 MOTIVATION

Discretization errors introduced by Newton-Cotes formulas and approximation errors of the VRI is widely studied topic. Etiene et al. [57] focused on the order of accuracy of the numerical solutions of the VRI when the number of samples increases. We try to devise closed form formulas for computing Eq. 3.10 from Section 3.3 to eliminate this kind of errors for our special case of occlusion maps (see Fig. 3.4).

5.3 NOTATION OF OPTICAL MODELS

To preserve the consistency with the existing literature, we adopt the Max's derivation of optical models for DVR [104]. For the sake of brevity, we omit simple cases when the participating medium is treated as a mass of cold and non-radiant perfectly black particles (i.e. absorption only model) or contrary, as an almost transparent hot tenuous gas (i.e. emission only model). Instead, we focus on the most common situation of participating medium containing particles that glow and also occlude passing radiant flux. Notice that even pure emission would be sufficient

for our purpose. This low albedo emission with absorption model for a single light ray is described by the differential equation

$$\frac{\mathrm{d}I}{\mathrm{d}t} = q(t) - \tau(t)I(t),$$ (5.1)

which relates the change in the rate of radiance I at the position t to the emissivity q of the particles located at the same infinitesimally small area and the radiance itself attenuated by the extinction coefficient τ, i.e. the rate that light is occluded. Solving the emission-absorption optical model in Eq. 5.1, describing the light transport through the participating medium, leads to the well known volume rendering integral (VRI)

$$I(D) = I_0 \exp\left(-\kappa \int_{t_0}^{D=t_n} \tau(t)\,\mathrm{d}t\right) + \int_{t_0}^{D} q(s) \exp\left(-\kappa \int_{s}^{D} \tau(t)\,\mathrm{d}t\right) \mathrm{d}s,$$ (5.2)

returning the radiance I of the volume at the exit point D aiming toward the camera [55]. Optical property κ is absorption coefficient, q is source term representing emission and the scalar field of densities is embodied in the function τ also known as an extinction coefficient and has units of number of particles per unit length at the parametric position t. Integration is taken from the entry point t_0 into the volume, to the exit point D toward the camera. This definite integral, more precisely Riemann integral as t is restricted to lie on the real line, cannot be solved analytically in general, instead, numerical methods are applied to find an approximation [55, 116]. In the following section, we will introduce the parametrization that will allow us do devise closed form solution for Eq. (5.2) for purposes of occlusion map evaluation.

5.4 INTERPOLATION OF THE SCALAR FIELD

Firstly, let us define the volume Ω which comprises of cells and vertices. To address individual cells, we define a set of all possible cell indices I as a triplet (i, j, k), i.e. the Cartesian product of cell indices ranges in all directions $\{0, \ldots, M-1\} \times \{0, \ldots, N-1\} \times \{0, \ldots, K-1\}$, where M, N, and K are numbers of vertices in give directions. To the each vertex we assign a value of interpolated quantity. This quantity may represent the density of a participating medium or, as in our case, the probability of a

Figure 5.1: The scheme for calculation of the line integral through i-th cell. Voxels are treated as vertices of cells, and cells fill the space between voxels

car's presence at each vertex location. From now, i will denote the entire cell indices triplet to simplify the notation.

Now, in accordance with Fig 5.1 we can introduce the following parametrization

$$\mathbf{u}(t) = \begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix} = (\mathbf{r}(t) - A) \oslash (G - A) = (O + \hat{\mathbf{d}}t - A) \otimes \mathbf{k}$$
$$= \underbrace{(O - A) \otimes \mathbf{k}}_{\mathbf{p}} + \underbrace{(\hat{\mathbf{d}} \otimes \mathbf{k})}_{\mathbf{s}} t = \mathbf{p} + \mathbf{s}t \tag{5.3}$$

which returns the relative position $\mathbf{u}$ of a sample against the cell's vertex A. Symbols $\oslash$ and $\otimes$ stand for element-wise division, resp. element-wise multiplication. We follow with the traditional trilinear interpolation

$$\gamma(t) = \beta_0(u(t), v(t))(1 - w(t)) + \beta_1(u(t), v(t))w(t), \tag{5.4}$$

where the subsequent functions are defined as follows

$$
\begin{aligned}
\beta_0(u,v) &= \alpha_{AB}(u)(1-v) + \alpha_{DC}(u)v, \\
\beta_1(u,v) &= \alpha_{EF}(u)(1-v) + \alpha_{HG}(u)v,
\end{aligned}
\tag{5.5}
$$

and

$$
\begin{aligned}
\alpha_{AB}(u) &= \rho(A)(1-u) + \rho(B)u, \\
\alpha_{DC}(u) &= \rho(D)(1-u) + \rho(C)u, \\
\alpha_{EF}(u) &= \rho(E)(1-u) + \rho(F)u, \\
\alpha_{HG}(u) &= \rho(H)(1-u) + \rho(G)u.
\end{aligned}
\tag{5.6}
$$

This parametrization leads to simple polynomial representation of tri-linear interpolation of the scalar function over the given cell

$$
\gamma(t) = \sum_{i=0}^{3} a_i t^i
\tag{5.7}
$$

where the individual coefficients a_i are listed in the Appendix, Table 7.1. Furthermore, for common shading algorithms the expression for normal at given point (i.e. $\nabla\gamma/||\nabla\gamma||$) is required. With the aim of chain rule

$$
\frac{\mathrm{d}}{\mathrm{d}t}\gamma(u(t),v(t),w(t)) = \frac{\partial\gamma}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\partial\gamma}{\partial v}\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{\partial\gamma}{\partial w}\frac{\mathrm{d}w}{\mathrm{d}t} = \nabla\gamma \cdot \underbrace{\left(\frac{\mathrm{d}u}{\mathrm{d}t}, \frac{\mathrm{d}v}{\mathrm{d}t}, \frac{\mathrm{d}w}{\mathrm{d}t}\right)}_{\mathbf{s}}
\tag{5.8}
$$

and Eq. 5.7, we can provide the first compounds of the $\nabla\gamma$ as follows

$$
\begin{aligned}
\frac{\partial}{\partial u}\gamma(t) = {}& (\rho(A) - \rho(B))(v(t) + w(t) - v(t)w(t) - 1) \\
& + (\rho(C) - \rho(D))(v(t) - v(t)w(t)) \\
& + (\rho(E) - \rho(F))(v(t)w(t) - w(t)) \\
& + (\rho(G) - \rho(H))v(t)w(t).
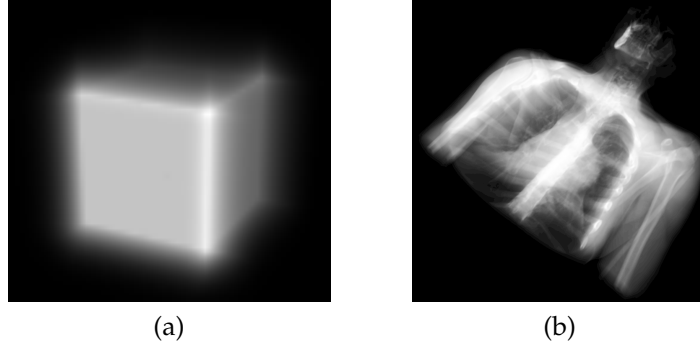\end{aligned}
\tag{5.9}
$$

65

Figure 5.2: (a) Example of a simple cube analytical ray tracing. (b) CT scan of a body visualized by the emission only model

The remaining two compounds are listed in Appendix, Eqs (7.1) and (7.2). Just for the sake of completeness of useful orders of derivatives, we also include the second derivative of γ

$$
\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\gamma) = \frac{\mathrm{d}}{\mathrm{d}t}(\nabla\gamma) \cdot \mathbf{s} = \big[& s_y\tau_1 + s_z\tau_2 + (p_ys_z + p_zs_y + 2s_ys_zt)\tau_3, \\
& s_x\tau_1 + s_z\tau_4 + (p_xs_z + p_zs_x + 2s_xs_zt)\tau_3, \\
& s_x\tau_2 + s_y\tau_4 + (p_xs_y + p_ys_x + 2s_xs_yt)\tau_3 \big] \cdot \mathbf{s}
\end{aligned}
\tag{5.10}
$$

where

$$
\begin{aligned}
\tau_1 &= \rho(A) - \rho(B) + \rho(C) - \rho(D), \\
\tau_2 &= \rho(A) - \rho(B) - \rho(E) + \rho(F), \\
\tau_3 &= \rho(E) - \rho(F) + \rho(G) - \rho(H) - \tau_1, \\
\tau_4 &= \rho(A) - \rho(D) - \rho(E) + \rho(H).
\end{aligned}
\tag{5.11}
$$

5.5 CONCLUSION

In this chapter, we have devised the analytical solution for Eq. 5.7 which appeared in the Eq. 3.10. This allows us to obtain the occlusion maps for parking spaces (see Fig. 3.4) in straight analytical way without utilizing numerical approximations. Devised formulas may be further applied in the more general area of DVR (see Fig. 5.2).

CONCLUSION

<div align="right">6</div>

The main objective of this dissertation thesis was to devise a new algorithm for estimating the occupancy of individual parking spaces from the static camera images of observed parking lot. One of the reasons that motivated our interest in this branch of static traffic surveillance was the collaboration with one car manufacturing company which was carried out between the years 2011 and 2012. This helped us to further particularize the requirements and constrains imposed on the desired parking guidance system. It is also worth to note that the parking lot occupation detection is an active research topic and has received attention from the computer vision community during the last decade. Tens of competitive approaches based on different principles was mentioned in the state of the art section but one feature they all have in common. They rely on carefully prepared data sets which are afterwards used during the training phase. This makes setup of those methods quite challenging and horrendous time consuming chore, especially for those who are not familiar with various caveats of particular method. To distinguish ourselves from this, and in reaction to it, we decided that our method should be parametric-based. Because of this, we tried to utilize as much of a priori known information about both the parking lot geometry and parked cars, to overcome the tedious process of training data set acquisition. Although this could be alleviated by semi-supervised learning techniques, the problem still prevails. Especially in the environment of outdoor parking lot, it could be very difficult to build a representative learning dataset covering all possible combinations of parking spaces states and surrounding conditions. As a consequence, the training process may introduce some level of unintentional bias.

As the main part of this work, we presented a new algorithm for the vision-based evaluation of parking lot utilization based on the analysis of the spatial arrangement of HOG features in the normalized image of a single parking space. The hypothesis about their expected arrangement in the case of occupied parking space is build upon the probabilistic car model which comprises the available data about the layout of the parking

lot and the camera system. The utilization of a priori known information allows us to remove the problematic learning phase, and as a result, the method does not require preparation of any training data set. The consistency of obtained features with an expected shape of potentially parked car is evaluated by the means of features prioritization through the adopted CFD technique that adds the flexibility to adapt the parked car appearance model to various possible configurations (e.g. car size, position, and yaw). The contextual constraints represented by the CRF ensure the spatial consistency of the final labeling. The integration of the self-learned parking lot background color model ensures the correctness of the inferred labeling in cases when the shape-prior is not reliable.

The experiments show that the algorithm performs well over the wide range of lighting conditions and the achieved F1 score was no lower than 98.1% in case of all parking rows. This result is comparable to the results of the most profound methods found in the literature. In our future work, we also plan to address an effective utilization of GPUs to reduce the overall latency of our parking lot surveillance system.

We also try to improve the HOG features which were used as a main descriptor for estimating the state of a parking space. This improvement renders the background subtraction to be significantly less vulnerable to the most common camera related phenomena like jitter and automatic exposure control.

In conjunction with the probabilistic car model, we have devised analytical solution to the related line integral. This result is directly applicable in the area of direct volume rendering.

At the time of writing, the here presented methods were published in [1, 2, 3, 4] and were provably cited in more than 20 conference papers, technical reports, journal articles, or dissertation thesis.

APPENDIX

7.1 PARKING LOT SCHEME
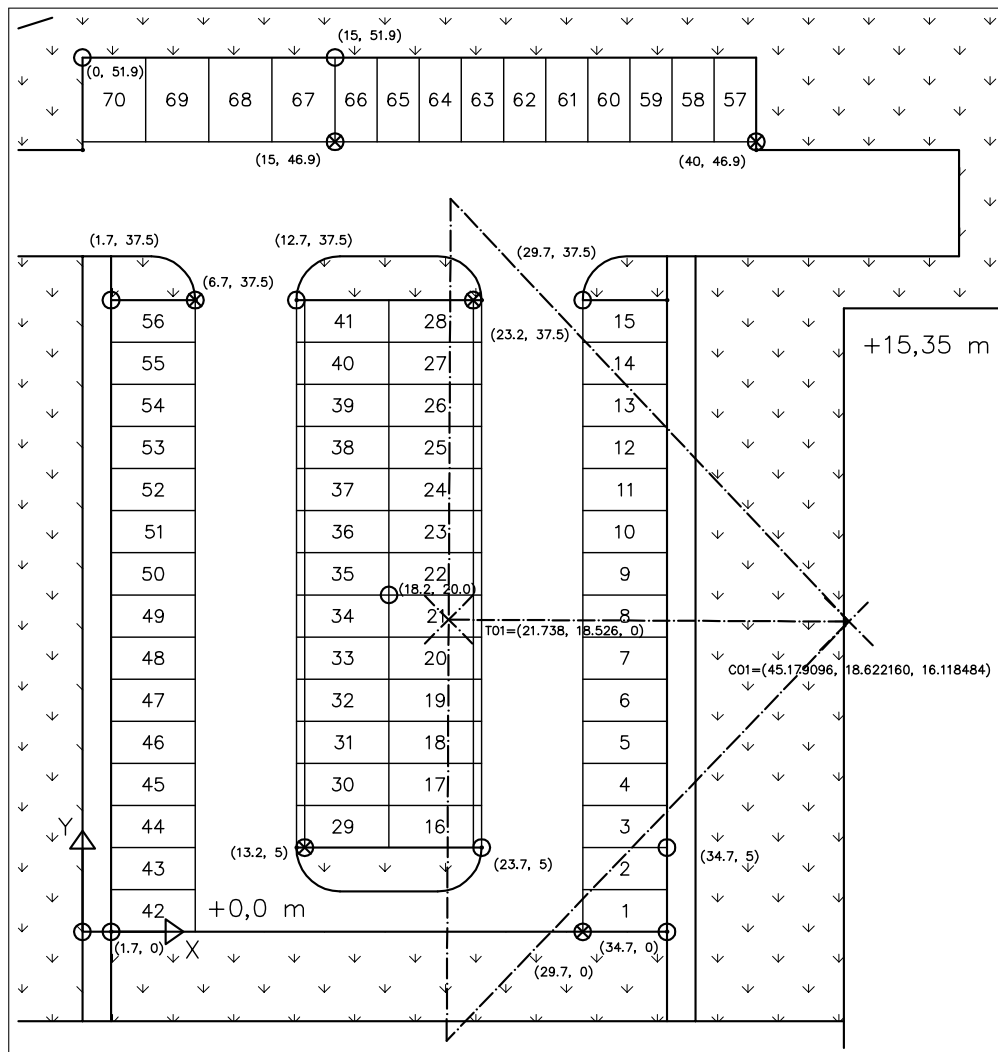
Figure 7.1: Scheme shows the parking lot layout, the camera position, and the calibration points

7.2 PARKING LOT OCCUPANCY

Figure 7.2: Parking lot with highlighted states of parking places as obtained from our method with both contextual constraints and background color model

7.3 DIRECT VOLUME RENDERING

$$
\begin{aligned}
\frac{\partial}{\partial v}\gamma(t) = {} & (\rho(A) - \rho(D))(u(t) + w(t) - u(t)w(t) - 1) \\
& + (\rho(B) - \rho(C))(u(t)w(t) - u(t)) \\
& + (\rho(G) - \rho(F))(u(t)w(t)) \\
& + (\rho(H) - \rho(E))(w(t) - u(t)w(t)) \, .
\end{aligned}
\tag{7.1}
$$

$$
\begin{aligned}
\frac{\partial}{\partial w}\gamma(t) = {} & (\rho(A) - \rho(E))(u(t) + v(t) - u(t)v(t) - 1) \\
& + (\rho(B) - \rho(F))(u(t)v(t) - u(t)) \\
& + (\rho(G) - \rho(C))(u(t)v(t)) \\
& + (\rho(D) - \rho(H))(u(t)v(t) - v(t)) \, .
\end{aligned}
\tag{7.2}
$$

Table 7.1: Coefficients of the cubic polynomial representing the trilinear interpolation of scalar values inside the cell. Notice that these are constants for the given ray and cell

$i = 0$	$\tau_{01} = p_x p_y \qquad \tau_{02} = p_x p_z \qquad \tau_{03} = p_y p_z \qquad \tau_{04} = \tau_{01} p_z$ $\begin{aligned} a_0 \ =\ & \rho(E)(\tau_{04} - \tau_{02} - \tau_{03} + p_z) \\ &+ \rho(F)(\tau_{02} - \tau_{04}) \\ &+ \rho(G)(\tau_{04}) \\ &+ \rho(H)(\tau_{03} - \tau_{04}) \\ &+ \rho(A)(\tau_{01} + \tau_{02} + \tau_{03} - \tau_{04} - p_x - p_y - p_z + 1) \\ &+ \rho(B)(\tau_{04} - \tau_{01} - \tau_{02} + p_x) \\ &+ \rho(C)(\tau_{01} - \tau_{04}) \\ &+ \rho(D)(\tau_{04} - \tau_{01} - \tau_{03} + p_y) \end{aligned}$
$i = 1$	$\begin{aligned} \tau_{05} &= \tau_{01} s_z + \tau_{02} s_y + s_x \tau_{03} \qquad \tau_{08} = p_x s_z \qquad \tau_{09} = s_x p_z \\ \tau_{10} &= p_y s_z \qquad \tau_{11} = s_y p_z \qquad \tau_{12} = p_x s_y \qquad \tau_{13} = s_x p_y \end{aligned}$ $\begin{aligned} a_1 \ =\ & \rho(E)(\tau_{05} - \tau_{08} - \tau_{09} - \tau_{10} - \tau_{11} + s.z) \\ &+ \rho(F)(\tau_{09} + \tau_{08} - \tau_{05}) \\ &+ \rho(G)\tau_{05} \\ &+ \rho(H)(\tau_{11} + \tau_{10} - \tau_{05}) \\ &+ \rho(A)(\tau_{12} - \tau_{05} + \tau_{08} + \tau_{13} + \tau_{09} + \tau_{10} + \tau_{11} \\ &\quad - s.x - s.y - s.z) \\ &+ \rho(B)(\tau_{05} - \tau_{12} - \tau_{08} - \tau_{13} - \tau_{09} + s.x) \\ &+ \rho(C)(\tau_{13} + \tau_{12} - \tau_{05}) \\ &+ \rho(D)(\tau_{05} - \tau_{12} - \tau_{13} - \tau_{10} - \tau_{11} + s.y) \end{aligned}$
$i = 2$	$\begin{aligned} \tau_{14} &= s_x s_z \qquad \tau_{15} = s_y s_z \qquad \tau_{16} = s_x s_y \\ \tau_{17} &= \tau_{14} p_y + p_x \tau_{15} + \tau_{16} p_z \end{aligned}$ $\begin{aligned} a_2 \ =\ & \rho(E)(\tau_{17} - \tau_{14} - \tau_{15}) \\ &+ \rho(F)(\tau_{14} - \tau_{17}) \\ &+ \rho(G)\tau_{17} \\ &+ \rho(H)(\tau_{15} - \tau_{17}) \\ &+ \rho(A)(\tau_{15} + \tau_{16} + \tau_{14} - \tau_{17}) \\ &+ \rho(B)(\tau_{17} - \tau_{16} - \tau_{14}) \\ &+ \rho(C)(\tau_{16} - \tau_{17}) \\ &+ \rho(D)(\tau_{17} - \tau_{16} - \tau_{15}) \end{aligned}$
$i = 3$	$\begin{aligned} a_3 \ =\ & s_x s_y s_z (\rho(E) - \rho(F) + \rho(G) - \rho(H) \\ &- \rho(A) + \rho(B) - \rho(C) + \rho(D)) \end{aligned}$

7.4 AUTHOR'S BIBLIOGRAPHY

LIST OF RELATED PUBLICATIONS

ISI Web of Knowledge

[1] T. Fabián. An algorithm for parking lot occupation detection. In *Proceedings of the International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, pages 165–170, Rožnov pod Radhoštěm, Czech Republic, June 2008.

[2] T. Fabián. Mixture of gaussians exploiting histograms of oriented gradients for background subtraction. In *Proceedings of the 6th International Symposium on Visual Computing (ISVC)*, pages 716–725, Las Vegas, Nevada, USA, November 2010.

[3] T. Fabián. A vision-based algorithm for parking lot utilization evaluation using conditional random fields. In *Proceedings of the 9th International Symposium on Visual Computing (ISVC)*, pages 222–233, Rethymnon, Crete, Greece, July 2013.

SciVerse Scopus

[4] T. Fabián. Parking lot occupancy detection using computational fluid dynamics. In *Proceedings of the 8th International Conference on Computer Recognition Systems (CORES)*, pages 733–742, Milkow, Poland, May 2013.

LIST OF UNRELATED PUBLICATIONS

ISI Web of Knowledge

[5] E. Sojka, J. Gaura, T. Fabián, and M. Krumnikl. Active contours without edges and without reinitialisation. In *Proceedings of the World Congress on Nature and Biologically Inspired Computing (NABIC)*, pages 914–917, Coimbatore, India, December 2009.

[6] E. Sojka, J. Gaura, Š. Šrubař, T. Fabián, and M. Krumnikl. Blurring mean-shift with a restricted data-set modification for applications in image processing. In *Proceedings of the 6th International Symposium on Visual Computing (ISVC)*, pages 310–319, Las Vegas, Nevada, USA, November 2010.

[7] E. Sojka, J. Gaura, T. Fabián, and M. Krumnikl. Fast mean shift algorithm based on discretisation and interpolation. In *Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS)*, pages 402–413, Sydney, Australia, December 2010.

[8] L. Ličev and T. Fabián. Development of methods for the processing of medical images using genetic algorithms. In *Proceedings of the 11th International Multidisciplinary Scientific GeoConference (SGEM)*, pages 495–502, Sofia, Bulgaria, June 2011.

[9] L. Ličev, T. Fabián, and J. Král. Image preprocessing and adjustments for photogrammetric measurement. In *Proceedings of the 11th International Multidisciplinary Scientific GeoConference (SGEM)*, pages 511–518, Sofia, Bulgaria, June 2011.

[10] L. Ličev, L. Skanderová, and T. Fabián. Analysis of mining images using active contours and self organizing migration algorithm. In *Proceedings of the 13th International Multidisciplinary Scientific GeoConference (SGEM)*, pages 3–10, Sofia, Bulgaria, June 2013.

[11] L. Ličev, M. Krumnikl, L. Skanderova, and T. Fabián. Design and implementation of the workplace for imaging and afterwards measurement of the plate in the area of the carotic artery. In *Proceedings of the 13th International Multidisciplinary Scientific GeoConference (SGEM)*, pages 16–22, Sofia, Bulgaria, June 2013.

SciVerse Scopus

[12] T. Fabián, J. Gaura, and P. Kotas. An algorithm for iris extraction. In *Proceedings of the 2nd International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 464–468, Paris, France, July 2010.

Journals with impact factor

[13] L. Ličev, R. Farana, T. Fabián, and J. Král. Image transformation and calibration using system fotom. *Acta Montanistica Slovaca (AMS)*, 16(1):1–7, 2011.
IF 0.053, 5-year IF 0.179

[14] L. Ličev, M. Babiuch, T. Fabián, and R. Farana. Development of methods for the processing of mining images using genetic algorithms. *Acta Montanistica Slovaca (AMS)*, 17(3):218–223, 2012.
IF 0.053, 5-year IF 0.179

Journals indexed by SciVerse Scopus

[15] T. Fabián, J. Gaura, and P. Kotas. Iris localization and extraction algorithm from an eye image. *International Journal of Signal and Imaging Systems Engineering (IJSISE)*, 5(2):69–77, 2012.
SJR 0.143

Journals with ISSN

[16] L. Ličev and T. Fabián. Application of genetic algorithms in the area of medical images segmentation. *International Journal on Information Technologies and Security*, 3(2):13–22, 2011.

Book chapters

[17] L. Ličev, I. Zelinka, and T. Fabián. Search and implementation of optimization algorithms in analysis of ultrasonic pictures in neurology. In I. Zelinka, V. Snášel, and A. Abraham, editors, *Handbook of Optimization*, pages 575–595. Springer, 2013.

PhD workshops

[18] T. Fabián and J. Gaura. Tracking moving objects in video traffic surveillance with classification by self-organizing map. In *Pro-*

ceedings of the 6th PhD Workshop (WOFEX), pages 282–287, Ostrava, Czech Republic, September 2008.

[19] T. Fabián and J. Gaura. Parallel implementation of recursive background modeling technique in cuda for tracking moving objects in video traffic surveillance. In *Proceedings of the 4th Doctoral Workshop on Mathematical and Engineering Methods in Computer Science (MEMICS)*, pages 29–36, Znojmo, Czech Republic, November 2008.

[20] T. Fabián. A comparative study of numerical methods for ray-nurbs intersection test. In *Proceedings of the 7th PhD Workshop (WOFEX)*, pages 257–262, Ostrava, Czech Republic, September 2009.

[21] T. Fabián. Background modeling and subtraction by local gradients. In *Proceedings of the 8th PhD Workshop (WOFEX)*, pages 230–235, Ostrava, Czech Republic, September 2010.

[22] T. Fabián, J. Gaura, and P. Kotas. Iris segmentation in eye image. In *Proceedings of the 9th PhD Workshop (WOFEX)*, pages 337–342, Ostrava, Czech Republic, September 2011.

[23] V. Ablavsky and S. Sclaroff. Layered graphical models for tracking partially occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1758–1775, September 2011. ISSN 0162-8828.

[24] V. Ablavsky, A. Thangali, and S. Sclaroff. Layered graphical models for tracking partially occluded objects. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Anchorage, Alaska, USA, June 2008.

[25] B. Antić, J. O. N. Castaneda, D. Ćulibrk, A. Pižurica, V. Crnojević, and W. Philips. Robust detection and tracking of moving objects in traffic video surveillance. In *Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS)*, pages 494–505, Bordeaux, France, September 2009.

[26] B. Antić, J. O. N. o. Castaneda, D. Ćulibrk, A. Pižurica, V. Crnojević, and W. Philips. Robust detection and tracking of moving objects in traffic video surveillance. In *Proceedings of the Advanced Concepts for Intelligent Vision Systems (ACIVS)*, pages 494–505, Bordeaux, France, September 2009.

[27] M. Balcilar and A. C. Sonmez. The effect of color space and block size on foreground detection. In *Proceedings of the 21st Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, Haspolat-Nikosia, North Cyprus, Turkey, April 2013.

[28] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, pages 404–417, Graz, Austria, May 2006.

[29] J. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, International Computer Science Institute, Berkeley, California, USA, 1998.

[30] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, International Computer Science Institute, Berkeley, California, USA, April 1998.

[31] D. Bong, K. Ting, and K. Lai. Integrated approach in the design of car park occupancy information system (COINS). *IAENG International Journal of Computer Science*, 35(1):7–14, 2009. ISSN 1819-9224.

[32] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR)*, pages 401–408, Amsterdam, Netherlands, July 2007.

[33] T. Bouwmans. Recent advanced statistical background modeling for foreground detection - a systematic survey. *Recent Patents on Computer Science*, 4(3):147–176, 2011. ISSN 1874-4796.

[34] C. Bravo, N. Sanchez, N. García, and J. M. M. García. Outdoor vacant parking space detector for improving mobility in smart cities. In *Proceedings of the 16th Portuguese Conference on Artificial Intelligence (EPIA)*, pages 30–41, Angra do Heroísmo, Azores, Portugal, September 2013.

[35] D. C. Brown. Decentering distortion of lenses. *Photometric Engineering*, 32(3):444–462, 1966.

[36] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the 8th European Conference on Computer Vision (ECCV)*, pages 25–36, Prague, Czech Republic, May 2004.

[37] M. Calonder, V. Lepetit, and P. Fua. BRIEF: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision (ECCV)*, pages 404–417, Heraklion, Crete, September 2010.

[38] A. Castano, A. Fukunaga, J. Biesiadecki, L. Neakrase, P. Whelley, R. Greeley, M. Lemmon, R. Castano, and S. Chien. Automatic detection of dust devils and clouds on Mars. *Machine Vision and Applications*, 19(5-6):467–482, 2008. ISSN 0932-8092.

[39] V. Chandrasekhar, G. Takacs, D. M. Chen, S. S. Tsai, R. Grzeszczuk, and B. Girod. CHoG: compressed histogram of gradients a low bit-rate feature descriptor. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2504–2511, Fontainebleau Resort, Miami Beach, Florida, USA, June 2009.

[40] R. Chellappa, X. Zhang, P. Burlina, C. Lin, Q. Zheng, L. Davis, and A. Rosenfeld. An integrated system for site model supported monitoring of transportation activities in aerial images. In *Proceedings of the DARPA Image Understanding Workshop*, pages 275–304, Palm Springs, California, USA, 1996.

[41] L.-C. Chen, J.-W. Hsieh, W.-R. Lai, C.-X. Wu, and S.-Y. Chen. Vision-based vehicle surveillance and parking lot management using multiple cameras. In *Proceedings of the 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 631–634, Darmstadt, Germany, October 2010.

[42] S.-C. S. Cheung and C. Kamath. Robust techniques for background subtraction in urban traffic. In *Proceedings of the Visual Communications and Image Processing (VCIP)*, pages 881–892, San Jose, California, USA, January 2004.

[43] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25:1337–1342, 2003.

[44] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, San Diego, California, USA, June 2005.

[45] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, San Diego, California, USA, June 2005.

[46] P. Dalka, D. Ellwart, and G. Szwoch. Camera orientation-independent parking events detection. In *Proceedings of the 12th*

International Workshop on Image Analysis for Multimedia Interactive Services, pages 13–22, Delft, Netherlands, April 2011.

[47] P. Dalka, G. Szwoch, and A. Ciarkowski. Distributed framework for visual event detection in parking lot area. In *Proceedings of the 4th International Conference on Multimedia Communications, Services and Security (MCSS)*, pages 37–45, Kraków, Poland, June 2011.

[48] N. Dan. Parking management system and method, July 2003. URL https://www.google.com/patents/US20030144890. US Patent App. 10/066,215.

[49] H. Dee and S. Velastin. How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, 19 (5):329–343, 2008. ISSN 0932-8092.

[50] D. Delibaltov, W. Wu, R. P. Loce, and E. A. Bernal. Parking lot occupancy determination from lamp-post camera images. In *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC)*, pages 2387–2392, Hague, Netherlands, October 2013.

[51] R. A. Drebin, L. Carpenter, and P. Hanrahan. Volume rendering. *SIGGRAPH Computer Graphics*, 22(4):65–74, 1988. ISSN 0097-8930.

[52] M. S. Drew. Recovery of chromaticity image free from shadows via illumination invariance. In *Proceedings of the IEEE Workshop on Color and Photometric Methods in Computer Vision (part of ICCV)*, pages 32–39, Nice, France, October 2003.

[53] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proceedings of the 6th European Conference on Computer Vision (ECCV)*, pages 751–767, Dublin, Ireland, June 2000.

[54] A. Elgammal, R. Duraiswami, D. Harwood, L. S. Davis, R. Duraiswami, and D. Harwood. Background and foreground modeling using nonparametric kernel density for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163, July 2002. ISSN 0018-9219.

[55] K. Engel, M. Hadwiger, J. Kniss, C. Rezk-Salama, and D. Weiskopf. *Real-Time Volume Graphics*. A K Peters, 1st edition, 2006. ISBN 978-1568812663.

[56] M. Enzweiler and D. M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(12):2179–2195, 2009. ISSN 0162-8828.

[57] T. Etiene, R. M. Kirby, and C. T. Silva. A study of discretization errors in volume rendering integral approximations. In *Proceedings of the EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, pages 92–99, Leipzig, Germany, June 2013.

[58] S. M. Faheem, G. Khan, M. Rahman, and H. Zafa. A survey of intelligent car parking system. *Journal of Applied Research and Technology*, 11(5):714–726, 2013. ISSN 1665-6423.

[59] P. F. Felzenszwalb and D. R. Huttenlocher. Efficient belief propagation for early vision. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 261–268, Washington, DC, USA, June 2004.

[60] G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, pages 823–836, Copenhagen, Denmark, May 2002.

[61] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *Proceedings of the 8th European Conference on Computer Vision (ECCV)*, pages 582–595, Prague, Czech Republic, May 2004.

[62] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1): 35–57, October 2009. ISSN 0920-5691.

[63] G. L. Foresti, C. Micheloni, and L. Snidaro. Event classification for automatic visual-based surveillance of parking lots. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, pages 314–317, Cambridge, England, August 2004.

Bibliography

[64] J. Frank. Intrinsic images for shadow removal. Technical report, School of Computer Science, McGill University, Montreal, Quebec, Canada, December 2004.

[65] N. Friedman and S. Russell. Image segmentation in video sequences: A probabilistic approach. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence (UIA)*, pages 175–181, Providence, Rhode Island, USA, August 1997.

[66] L. M. Fuentes and S. A. Velastin. From tracking to advanced surveillance. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 121–124, Barcelona, Catalonia, Spain, September 2003.

[67] S. Funck, N. Möhler, and W. Oertel. Determining car-park occupancy from single images. In *Proceedings of the IEEE Symposium on Intelligent Vehicles (IV)*, pages 325–328, Parma, Italy, June 2004.

[68] V. Gouaillier and A.-E. Fleurant. Intelligent video surveillance: Promises and challenges. Technical report, Centre de recherche informatique de Montréal and Technopôle Defence and Security, Montréal, Québec, Kanada, April 2009.

[69] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. ISSN 0004-3702.

[70] C.-C. Huang and S.-J. Wang. A hierarchical bayesian generation framework for vacant parking space detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(12):1770–1785, December 2010. ISSN 1051-8215.

[71] C.-C. Huang, S.-J. Wang, Y.-J. Chang, and T. Chen. A bayesian hierarchical detection framework for parking space detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2097–2100, Las Vegas, Nevada, USA, March 2008.

[72] C.-C. Huang, Y.-S. Tai, and S.-J. Wang. Vacant parking space detection based on plane-based bayesian hierarchical framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(9):1598–1610, September 2013. ISSN 1051-8215.

[73] H. Ichihashi, K. Nagaura, A. Notsu, and K. Honda. Benchmarking parameterized fuzzy c-means classifier. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1137–1144, Jeju Island, Korea, August 2009.

[74] H. Ichihashi, A. Notsu, and K. Honda. Triplet of fcm classifiers. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1826–1833, Jeju Island, Korea, August 2009.

[75] H. Ichihashi, A. Notsu, K. Honda, T. Katada, and M. Fujiyoshi. Vacant parking space detector for outdoor parking lot by using surveillance camera and fcm classifier. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 127–134, Jeju Island, Korea, August 2009.

[76] H. Ichihashi, T. Katada, M. Fujiyoshi, A. Notsu, and K. Honda. Improvement in the performance of camera based vehicle detector for parking lot. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1–7, Barcelona, Spain, July 2010.

[77] M. Y. I. Idris, Y. Leng, E. M. Tamil, N. M. Noor, and Z. Razak. Car park system: A review of smart parking system and its technology. *Information Technology Journal*, 8(2):101–113, 2009. ISSN 1812-5638.

[78] IEEE. *IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries 610*. Institute of Electrical and Electronics Engineers, 1990. ISBN 978-1559370790.

[79] T. H. Im, I. K. Eom, and Y. S. Kim. Wavelet-based moving object segmentation using background registration technique. In *Proceedings of the 9th IASTED International Conference on Signal and Image Processing (SIP)*, pages 84–88, Honolulu, Hawaii, USA, August 2007.

[80] R. Jain and H. Nagel. On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1(2):206–214, 1979. ISSN 0162-8828.

[81] J. Jermsurawong, U. Ahsan, A. Haidar, H. Dong, and N. Mavridis. One-day long statistical analysis of parking demand by using single-camera vacancy detection. *Journal of Transportation Systems*

Engineering and Information Technology, 14(2):33–44, May 2014. ISSN 1570-6672.

[82] P.-M. Jodoin, J. Konrad, V. Saligrama, and V. Veilleux-Gaboury. Motion detection with an unstable camera. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 229–232, San Diego, California, USA, October 2008.

[83] H. Jung, D. Kim, P. Yoon, and J. Kim. Stereo vision based localization of free parking site. In *Proceedings of the 11th International Conference on Computer Analysis of Images and Patterns (CAIP)*, pages 231–239, Paris, France, September 2005.

[84] H. G. Jung, C. G. Choi, P. J. Yoon, and J. Kim. Semi-automatic parking system recognizing parking lot markings. In *Proceedings of the 8th International Symposium on Advanced Vehicle Control (AVEC)*, pages 947–952, Taipei, Taiwan, August 2006.

[85] H. G. Jung, D. S. Kim, P. J. Yoon, and J. Kim. Two-touch type parking slot marking recognition for target parking position designation. In *Proceedings of the IEEE Symposium on Intelligent Vehicles (IV)*, pages 1161–1166, Eindhoven, Netherlands, June 2008.

[86] N. Kaempchen, U. Franke, and R. Ott. Stereo vision based pose estimation of parking lots using 3d vehicle models. In *Proceedings of the IEEE Symposium on Intelligent Vehicles (IV)*, pages 459–464, Versailles, France, June 2002.

[87] Z. Kalal, K. Mikolajczyk, and J. Matas. Face-TLD: Tracking-Learning-Detection Applied to Faces. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 3789–3792, Hong Kong, September 2010.

[88] F. Kalaycilar. An object recognition framework using contextual interactions among objects. Master's thesis, Bilkent University, Ankara, Turkey, 2009.

[89] H. Kang, D. Kim, and S. Y. Bang. Real-time multiple people tracking using competitive condensation. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 325–328, Rochester, New York, USA, September 2002.

[90] K. P. Karmann and A. von Brandt. Moving object recognition using and adaptive background memory. In *Time-Varying Image Process Moving Object Recognition*, volume 2, pages 289–307. Elsevier, Amsterdam, The Netherlands, 1990.

[91] Z. Khan, T. Balch, and F. Dellaert. Mcmc-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(11): 1805–1819, 2005. ISSN 0162-8828.

[92] T. H. Kim, I. K. Eom, and Y. S. Kim. Wavelet-based moving object segmentation using background registration technique. In *Proceedings of the International Conference on Signal and Image Processing (SIP)*, pages 84–88, Honolulu, Hawaii, USA, August 2007.

[93] S. Kumar and M. Hebert. Discriminative fields for modeling spatial dependencies in natural images. In *NIPS*. MIT Press, 2003.

[94] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, pages 282–289, San Francisco, California, USA, June 2001.

[95] N. Lazarevic-McManus, J. Renno, D. Makris, and G. A. Jones. An object-based comparative methodology for motion detection based on the f-measure. *Computer Vision and Image Understanding*, 111(1): 74–85, 2008. ISSN 1077-3142.

[96] S. Lee, D. Yoon, and A. Ghosh. Intelligent parking lot application using wireless sensor networks. In *Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS)*, pages 48–57, Irvine, California, USA, May 2008.

[97] M. Levoy. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(3):29–37, 1988. ISSN 0272-1716.

[98] S.-F. Lin, Y.-Y. Chen, and S.-C. Liu. A vision-based parking lot management system. In *Proceedings of the International Conference on Systems, Man and Cybernetics (SMC)*, pages 2897–2902, Taipei, Taiwan, October 2006.

[99] B. P. L. Lo and S. A. Velastin. Automatic congestion detection system for underground platforms. In *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP)*, pages 158–161, Kowloon Shangri-La, Hong Kong, May 2001.

[100] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004. ISSN 0920-5691.

[101] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 121–130, Washington, DC, USA, April 1981.

[102] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679, Vancouver, Canada, August 1981.

[103] I. Masmoudi, A. Wali, and A. M. Alimi. Parking spaces modelling for inter spaces occlusion handling. In *Proceedings of the 22nd International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pages 1–6, Pilsen, Czech Republic, June 2014.

[104] N. Max. Optical models for direct volume rendering, 1995.

[105] N. J. B. McFarlane and C. P. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 8(3):187–193, 1995.

[106] P. Meer. Robust computer vision: An interdisciplinary challenge. *Computer Vision and Image Understanding*, 78:1–7, 2000. ISSN 1077-3142.

[107] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 27(10):1615–1630, 2005. ISSN 0162-8828.

[108] L. E. Y. Mimbela and L. A. Klein. A summary of vehicle detection and surveillance technologies used in intelligent transportation systems. Technical report, Southwest Technology Development Institute at New Mexico State University, Las Cruces, New Mexico, USA, August 2007.

[109] K. C. Mouskos, M. Boile, and M. Boile. Technical solutions to overcrowded park and ride facilities. Technical report, The City College of New York, New York, New York, USA, May 2007.

[110] A. Nallamuthu and S. Lokala. Vision based parking space classification. Technical report, Clemson University, Clemson, South Carolina, USA, December 2008.

[111] J. Pan and B. Hu. Robust occlusion handling in object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Minneapolis, Minnesota, USA, June 2007.

[112] C. Papageorgiou and T. Poggio. A trainable system for object detection. *International Journal of Computer Vision (IJCV)*, 38(1):15–33, 2000. ISSN 0920-5691.

[113] M. A. Z. Pisheh and A. Sheikhi. Detection and compensation of image sequence jitter due to an unstable ccd camera for video tracking of a moving target. In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, pages 258–261, Thessaloniki, Greece, September 2004.

[114] P. W. Power and J. A. Schoonees. Understanding background mixture models for foreground segmentation. In *Proceedings of the Image and Vision Computing*, pages 267–271, Auckland, New Zealand, November 2002.

[115] P. W. Power and J. A. Schoonees. Understanding background mixture models for foreground segmentation. In *Proceedings of the Image and Vision Computing*, pages 267–271, Auckland, New Zealand, November 2002.

Bibliography

[116] B. Preim and D. Bartz. *Visualization in Medicine: Theory, Algorithms, and Applications*. Morgan Kaufmann, 1st edition, 2007. ISBN 978-0123705969.

[117] A. Rabinovich and S. Belongie. Scenes vs. objects: a comparative study of two approaches to context based recognition. In *Proceedings of the International Workshop on Visual Scene Understanding (ViSU)*, pages 92–99, Miami, Florida, USA, June 2009.

[118] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *Proceedings of the 11th International Conference on Computer Vision (ICCV)*, pages 1–8, Rio de Janeiro, Brazil, October 2007.

[119] P. Remagnino, A. Baumberg, T. Grove, D. Hogg, T. N. Tan, A. D. Worrall, and K. D. Baker. An integrated traffic and pedestrian model-based vision system. In *Proceedings of the British Machine Vision Conference (BMVC)*, Essex, UK, 1997.

[120] A. Sanin, C. Sanderson, and B. C. Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Journal of Pattern Recognition*, 45(4):1684–1695, April 2012. ISSN 0031-3203.

[121] R. J. L. Sastre, P. G. Jiménez, F. J. Acevedo, and S. M. Bascón. Computer algebra algorithms applied to computer vision in a parking management system. In *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE)*, pages 1675–1680, Vigo, Spain, June 2007.

[122] Y.-W. Seo. Augmenting cartographic resources and assessing roadway state for vehicle navigation. Technical report, The Robotics Institute Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, April 2012.

[123] Y.-W. Seo and C. Urmson. Utilizing prior information to enhance self-supervised aerial image analysis for extracting parking lot structures. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 339–344, Saint Louis, Missouri, USA, October 2009.

[124] Y.-W. Seo, N. Ratliff, and C. Urmson. Self-supervised aerial image analysis for extracting parking lot structure. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence (IJCAI)*, pages 1837–1842, Pasadena, California, USA, July 2009.

[125] Y.-W. Seo, C. Urmson, D. Wettergreen, and J.-W. Lee. Augmenting cartographic resources for autonomous driving. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 13–22, Seattle, Washington, USA, November 2009.

[126] J. Stam. Stable fluids. In *Proceedings of the 26th Annual Conference on Computer Graphics (SIGGRAPH)*, pages 121–128, Los Angeles, California, USA, August 1999.

[127] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 246–252, Fort Collins, Colorado, USA, June 1999.

[128] L. D. Stefano, G. Neri, and E. Viarani. Analysis of pixel-level algorithms for video surveillance applications. In *Proceedings of the 11th International Conference on Image Analysis and Processing (ICIAP)*, pages 542–546, Palermo, Italy, September 2001.

[129] J. K. Suhr and H. G. Jung. Fully-automatic recognition of various parking slot markings in around view monitor (avm) image sequences. In *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1294–1299, Anchorage, Alaska, USA, September 2012.

[130] J. K. Suhr and H. G. Jung. Sensor fusion-based vacant parking slot detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):21–36, 2014. ISSN 1524-9050.

[131] J. K. Suhr, H. G. Jung, K. Bae, and J. Kim. Automatic free parking space detection by using motion stereo-based 3d reconstruction. *Machine Vision and Applications*, 21(2):163–176, February 2010. ISSN 0932-8092.

[132] N. True. Vacant parking space detection in static images. Technical report, University of California, San Diego, California, USA, March 2007.

[133] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, pages 511–518, Kauai, Hawaii, USA, December 2001.

[134] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV)*, pages 734–741, Nice, France, October 2003.

[135] S. Wanayuth, A. Ohya, and T. Tsubouchi. Parking place inspection system utilizing a mobile robot with a laser range finder. In *Proceedings of the IEEE/SICE International Symposium on System Integration (SII)*, pages 55–60, Fukuoka, Japan, December 2012.

[136] S. Wanayuth, A. Ohya, and T. Tsubouchi. Inside vehicle inspection system utilizing a mobile robot with lrf sensor. *Journal of Robotics and Mechatronics*, 26(1):59–67, February 2014. ISSN 0915-3942.

[137] X. Wang and A. R. Hanson. Parking lot analysis and visualization from aerial images. In *Proceedings of the 4th Workshop on Applications of Computer Vision (WACV)*, pages 36–41, Princeton, New Jersey, USA, October 1998.

[138] X. Wang, T. X. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. In *Proceedings of the 12th International Conference on Computer Vision (ICCV)*, pages 32–39, Kyoto, Japan, September 2009.

[139] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 19(7):780–785, 1997. ISSN 0162-8828.

[140] Q. Wu and Y. Zhang. Parking lots space detection. Technical report, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, December 2006.

[141] Q. Wu, C. Huang, S. yu Wang, W. chen Chiu, and T. Chen. Robust parking space detection considering inter-space correlation. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 659–662, Bejing, China, July 2007.

[142] K. Yamada and M. Mizuno. A vehicle parking detection method using image segmentation. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 84(10):25–34, October 2001. ISSN 1520-6440.

[143] B. Yang, P. Sharma, and R. Nevatia. Vehicle detection from low quality aerial lidar data. In *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV)*, pages 541–548, Snowbird, Utah, USA, January 2011.

[144] T. Yang, Q. Pan, J. Li, and S. Z. Li. Real-time multiple objects tracking with occlusion handling in dynamic scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 970–975, San Diego, California, USA, June 2005.

[145] S. yi Chien, S. yih Ma, and L.-G. Chen. Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7): 577–586, 2002. ISSN 1051-8215.

[146] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4):1–45, 2006. ISSN 0360-0300.

[147] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22 (11):1330–1334, 2000. ISSN 0162-8828.

[148] J. Zhong, L. Gu, J. Tan, and Y. Li. Estimating parameters of gmm based on split em. *Computer Engineering and Applications*, 48(34): 28–32, 2012. ISSN 1002-8331.