Vacant Parking Space Detection Based on Plane-based Bayesian Hierarchical Framework

Ching-Chun Huang, Member, IEEE, Yu-Shu Tai, and Sheng-Jyh Wang, Member, IEEE

Abstract—In this paper, we propose a vacant parking space detection system that operates day and night. In the daytime, the major challenges of the system include dramatic lighting variations, shadow effect, inter-object occlusion, and perspective distortion. In the nighttime, the major challenges include insufficient illumination and complicated lighting conditions. To overcome these problems, we propose a plane-based method which adopts a structural 3-D parking lot model consisting of plentiful planar surfaces. The plane-based 3-D scene model plays a key part in handling inter-object occlusion and perspective distortion. On the other hand, to alleviate the interference of unpredictable lighting changes and shadows, we propose a plane-based classification process. Moreover, by introducing a Bayesian hierarchical framework to integrate the 3-D model with the plane-based classification process, we systematically infer the parking status. Last, to overcome the insufficient illumination in the nighttime, we also introduce a preprocessing step to enhance image quality. The experimental results show that the proposed framework can achieve robust detection of vacant parking spaces in both daytime and nighttime.

Index Terms—Bayesian inference, histogram of oriented gradients, image classification, parking space detection.

I. INTRODUCTION

RECENTLY, video surveillance systems have become increasingly important in our daily life. With the noticeable progress of computer vision techniques, many video surveillance systems have been proposed to provide new kinds of intelligent functions, like object detection and tracking. Following the trend, vision-based systems for smart parking lot management have also attracted great attention in recent years. In general, these vision-based parking lot management systems can provide valuable information, like the location of vacant parking spaces, as well as some value-added services, like parking space guidance and vehicle finding. In this paper, we focus on a basic, yet crucial, function of vision-based parking lot management systems automatic detection of vacant parking spaces.

In Fig. 1, we show several parking lot images in our dataset. To robustly detect vacant parking spaces, we have to deal with a few challenges, including dramatic lighting variations, shadows cast on the scene, varying perspective distortion in the image, and inter-object occlusion among parked cars and the ground plane. Besides, insufficient illumination during the nighttime is another challenge. To overcome these problems, many novel methods have been proposed in the past. These methods can be roughly categorized into four major types: car-oriented methods, space-oriented methods, hybrid methods, and parking-lot-oriented methods.

Car-oriented methods [1]–[5] target car detection, and they determine the status of parking spaces based on the detection result. Space-oriented methods [6]–[13] model the appearance of the ground plane in advance. If the current appearance of a parking space is dissimilar to that model, they identify the parking space as occupied. Some hybrid methods [14]–[16], on the other hand, combine both space detection and car detection to find vacant parking spaces. These hybrid methods focus on the design of the fusion mechanism to achieve improved performance. Recently, unlike car-oriented or space-oriented methods which focus only on certain aspects of parking lots, parking-lot-oriented methods [17], [18] have been proposed to model the whole parking lot in unity and to integrate the 3-D scene model with the image observation for parking status inference.

For car-oriented methods, Tsai et al. [1] propose a global color-based model to efficiently detect vehicle candidates. In
their approach, a Bayesian classifier is trained to verify the detection of vehicles based on corners, edges, and wavelet features. In [2], Meij et al. propose a color-based texture segmentation process for vehicle detection based on color and texture features. Masaki [3] keeps tracking and recording the movement of vehicles in order to identify vacant parking spaces. On the other hand, many algorithms [4]-[5] adopt certain consistent texture features, such as histogram of oriented gradients (HOG) [19], to overcome lighting variations and geometric distortion. In general, these methods can achieve robust detection even under dramatic variations of lighting condition. However, for vacant space detection, these car-oriented methods do not take into account the inter-vehicle occlusion problem.

For space-oriented methods, the modeling of parking spaces is the key. Eigen-space representation [6] and many background modeling algorithms [7]-[9] provide pixel-based methods to provide ground models that can adapt to lighting variations. However, these pixel-based space modeling methods are usually sensitive to the shadows cast over the ground. To relieve the shadow effect, some texture-based methods assume that a vacant parking space possesses homogeneous appearance. Hence, they design certain measure of homogeneity to detect vacant parking spaces. For example, Yamada et al. [10] design a homogeneity measure by calculating the area of fragmental segments; Lee et al. [11] suggest an entropy-based homogeneity metric; and Fabian [12] uses a segment-based homogeneity measure similar to that in [10]. However, due to perspective distortion, a distant parking space may only occupy a small region in the captured image. This usually leads to unstable homogeneity measurement. To overcome the perspective distortion problem, López-Sastre et al. [13] suggest a method to rectify the perspective distortion and they use a Gabor filter bank to derive the homogeneity feature for vacant parking space detection. Basically, these space-oriented methods still suffer from the inter-object occlusion problem, which occurs when a parking space is partially or fully occluded by a car at an adjacent parking space. Some researchers adopt hybrid methods to detect vacant parking spaces. For example, Dan [14] trains a general support vector machine (SVM) classifier to differentiate car regions from space regions by using image features made of the color vectors inside the parking space. However, this method cannot properly handle the inter-occlusion problem. To overcome the occlusion problem, Wu et al. [15] propose a method to group three neighboring spaces as a unit and they define the color histogram of the three-space unit as the feature in their SVM classifier. Even though these hybrid methods have considered both car model and space model, the classification performance of their algorithms is still affected by the environmental variations. In general, the lighting changes may dramatically degrade the detection accuracy. On the other hand, in [16], the authors propose an efficient method to combine static and dynamic information for vacant parking space detection. To extract static information, a histogram classification process is used to detect pavement regions while an edge counting process is used to identify vehicle regions. To extract dynamic information, they use blob analysis to track moving vehicles.

Table I lists all the comparisons of vacant space detection algorithms for five types of challenges, perspective distortion (PD), inter-object occlusion (IO), shadow effect (SE), lighting variations (LV), and insufficient illumination at night (IN). Meaning of symbols: X: Not Good Enough; ∆: Fair; O: Good.

In order to alleviate the inter-occlusion problem, however, their camera usually needs to be placed at a very high altitude. Rather than focusing on the detection of individual cars or parking spaces, parking-lot-oriented methods model the geometric structure of the whole parking lot in order to properly handle the inter-occlusion situations. In [17]-[18], Huang et al. propose a Bayesian hierarchical framework (BHF) to integrate the 3-D scene knowledge and the classification of image pixels into a three-layer hierarchical framework. The structural scene properties of a parking lot, together with the pixel-based car model and parking space model, are well utilized to improve the performance of vacant space detection. Moreover, to conquer the variations of lighting condition and the shadows cast on the scene, Huang et al.’s method assumes that the parking lot scene is uniformly lighted by sunlight and they have made a lot of effort to dynamically estimate the lighting condition. However, although their method can produce robust detection results in the daytime, it fails in the nighttime due to the complicated lighting condition at night. Actually, so far as we know, very few systems have ever discussed the vacant parking space detection problem in the nighttime.

For the sake of clarification, we summarize in Table I the comparisons of several algorithms for vacant parking space detection. As indicated in this table, none of these existing methods can handle all the five types of challenges, including perspective distortion, inter-object occlusion, shadow effect, lighting variations, and insufficient illumination at night. In this paper, a new parking-lot-oriented method is presented to deal with all these challenges. In the proposed method, we further improve the Bayesian hierarchical framework (BHF) in [18] to achieve robust detection of vacant parking spaces in both daytime and nighttime. In our method, we model the whole parking lot as a 3-D

<table>
<thead>
<tr>
<th>Type</th>
<th>Method</th>
<th>PD</th>
<th>IO</th>
<th>SE</th>
<th>LV</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>Tao [1]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Space</td>
<td>Background Modeling [7]-[9]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Space</td>
<td>Yamada [10]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Space</td>
<td>Lee et al. [11]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Dan [14]</td>
<td>∆</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Wu [15]</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Blumer [18]</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Parking lot</td>
<td>Huang [17]</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Parking lot</td>
<td>Huang [18]</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Parking lot</td>
<td>Proposed Method</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
structure consisting of plentiful planar surfaces. A plane-based classification process using robust texture features is proposed to replace the pixel-based classification in [18]. Furthermore, by using a modified BHF framework for inference, we can systematically model the relation between the 3-D planar surfaces and their image appearance. The inter-vehicle occlusion is well modeled in the modified BHF framework and illumination-insensitive object textures are well used for robust parking space detection. Furthermore, by introducing a multi-exposure pre-process to enhance the captured image sequence, we can perform vacant parking space detection days and nights under a unified framework.

The rest of this paper is organized as follows. In Section II, we briefly introduce the overview of the proposed system. In Section III, we illustrate the preprocessing stage for generating high-dynamic-range images in the nighttime. In Section IV, we present the proposed plane-based BHF inference framework for vacant space detection. Experimental results and discussions are presented in Section V. Last, Section VI concludes this paper.

II. OVERVIEW OF THE PROPOSED METHOD

In order to develop a vacant parking space detection system that can work all day, we focus on two major issues. The first issue is about how to obtain well exposed images for inference. In an outdoor scene, the lighting condition may have dramatic changes. Those variations greatly affect the appearance of image features, such as edges or colors, especially in the nighttime. To deal with this issue, we adopt a pre-process to enhance the visibility of image contents. On the other hand, the second issue is about how to improve the performance of vacant parking space detection and how to speed up the system for practical applications. To deal with this issue, a plane-based BHF framework is proposed for vacant parking space detection. By decomposing a parking lot into many 3-D planar surfaces, we can effectively exploit the texture information for vacant parking space detection and well represent the patterns of inter-vehicle occlusion.

In Fig. 2, we show the workflow of the proposed method, which consists of a preprocessing step and a detection step. In the preprocessing step, we design a multi exposure system to capture images with different exposure settings. These images are then fused to obtain images with improved quality. In the detection step, a plane-based BHF inference framework is proposed. First, based on the proposed plane-based 3-D scene model, the normalized patches of interest, corresponding to the projection of 3-D surfaces onto the fused image, are identified. For each normalized patch, the histogram of oriented gradients (HOG) features are extracted and are further compressed via linear discriminant analysis (LDA) [20]. Finally, we use the proposed plane-based BHF framework to integrate 3-D scene information with plane-based classification results for the optimal inference of the status of the parking spaces. In the following sections, we will explain the details of the proposed system in steps.

III. PREPROCESSING STEP

When capturing images in a dark environment, the color and texture information degrades. The degradation of image features may dramatically deteriorate the performance of vacant parking space detection. Hence, a pre-processing stage is used in our system to enhance the quality of nighttime images. Up to now, plentiful methods have been proposed to enhance image contrast, like the Retinex-based algorithms in [21], [22], the histogram-equalization-based algorithms in [23]-[26], the Gray level grouping method in [27], [28], the discrete cosine transform DCT-based method in [29], the tone-mapping method in [30], and the Bayesian inference method in [31]. Although those methods can improve image quality impressively, some side-effects, like noise amplification and halo effects, may generate extra image features and harm the following detection process. Different from those approaches which are based on a single image, we enhance nighttime images based on multiple images under different exposure settings. In a dark environment, some image features, like colors or edges, may be missing if the exposure time is too short, as shown in Fig. 3(a). On the contrary, image color or intensity may get saturated if the exposure time is too long,
as shown in Fig. 3(c). The choice of exposure time is usually a trade-off. With the use of multiple images under different exposure settings, we are able to extract useful image features in both dark and bright areas. By fusing these images into a single image, we can obtain an image with improved details, as shown in Fig. 3(d).

To get multi-exposure images, we use the AXIS M114 IP camera which can adjust the exposure value (EV) during image capturing. By using the software development kit (SDK) provided by AXIS, we capture images from a short exposure period to a long exposure period in a cyclic manner with the period of \( N \) image frames, as illustrated in Fig. 4. In our system, the longest and shortest exposure period is 3 s and 0.33 s, respectively. By using the two-step exposure fusion method proposed in [32] to combine every \( N \) images, we get images of improved contrast.

### IV. Detection Step

#### A. Plane-Based Structure and Feature Extraction

In our system, we attempt to find a way to benefit from both car-oriented and space-oriented approaches. For car-oriented methods, they usually check a car area like that in Fig. 5(a); while for space-oriented methods, they check the ground area like that in Fig. 5(b). In our approach, we treat the parking spaces as a set of cuboids, as illustrated in Fig. 5(c). Each cuboid is composed of six patches, as illustrated in Fig. 5(d).

Based on the 3-D cuboid model, we represent the structure of parking by a set of 3-D planar surfaces, as shown in Fig. 5(e). By projecting those 3-D surfaces onto the image, we get image patches of parallelogram shape. These patches are to be used for the status inference of parking spaces.

Due to the perspective projection in image formation, image patches may appear to be quite different in shape and size. To overcome perspective distortion, we normalize each image patch into a rectangle, with \( R_l \) pixels in length and \( R_w \) pixels in width. After that, we extract features from the normalized patches. For feature extraction, we adopt the HOG feature proposed in [19], which is less affected by shadows and the changes of illumination. To extract HOG features, a normalized image patch is regularly segmented into non-overlapping cells, with each cell containing \( C_l \times C_w \) pixels. In total, there would be \( (R_l/C_l) \times (R_w/C_w) \) cells in each normalized patch. For each cell, a histogram of oriented gradients, as defined in [19], is built. Each histogram has \( H_b \) histogram bins. By combining the histograms of all cells in the normalized patch, we obtain the HOG feature. In our system, the parameters \( (R_l/C_l, R_w/C_w, H_b) \) are empirically chosen to be \((64 \times 32, 16 \times 16, 10)\). That is, each normalized patch contains eight cells and the dimensions of its HOG feature is 80.

In Fig. 6, we illustrate the processes of patch normalization and HOG feature extraction. As will be explained later, these high-dimensional HOG features will be converted into low-dimensional features via LDA so that the following inference process can be implemented in a more efficient way.

#### B. Patch Classification

In this section, we explain how to perform patch classification in the proposed plane-based BHF framework. As mentioned before, in our plane-based model, each parking space is approximated as a cuboid with six 3-D planar surfaces. We classify these surfaces into four different types: ground surface (G), side surface (S), front (or rear) surface (F), and top surface (T). Via perspective projection, these four types of planar surfaces are projected onto four types of image patches:
the notation TypeIndex, where Type each surface type. In the following paragraphs, we will use and the four different combinations of parking statuses for patches related to the four different types of planar surfaces "X" means "do not care." In total, there are 16 kinds of image In this table, "o" means "occupied," "v" means "vacant," and parking space and the most influential adjacent parking space. type according to the four status combinations of the present define the indices of the four subclasses for each surface on the other hand, we may either classify its image patterns into four sub-classes that relate to the four status combinations of the spaces "b" and "c," or into eight sub-classes that relate to the eight status combinations of "a," "b," and "c." In our experiments, for the sake of simplification, we choose the four-subclass classification for T-patches. In Table II, based on the illustration in Fig. 7, we further define the indices of the four subclasses for each surface type according to the four status combinations of the present parking space and the most influential adjacent parking space. In this table, "o" means "occupied," "v" means "vacant," and "X" means "do not care." In total, there are 16 kinds of image patches related to the four different types of planar surfaces and the four different combinations of parking statuses for each surface type. In the following paragraphs, we will use the notation TypePatch, where Type ∈ {T, G, S, F} and Index ∈ {1, 2, 3, 4}, to label these 16 kinds of image patches. The whole set of these 16 patch labels is denoted as L = {T−1, T−2, T−3, T−4, G−1, G−2, G−3, G−4, S−1, S−2, S−3, S−4, F−1, F−2, F−3, F−4}. In Fig. 8, we illustrate the 16 kinds of image patches for a parking space, together with some patch samples. Note that the front surface and the rear surface belong to the same surface type. Similarly, the surfaces on the two sides of the parking space belong to the same surface type. It can be observed in these samples that the image content inside a patch may reveal not only the information of the current parking space but also the information of the adjacent parking space. Moreover, for each surface type, the image contents for different combinations of parking statuses appear to be quite different. Hence, it would be possible for us to classify a given image patch into one of the four subclasses simply based on its image content. The classification result provides evidence to support not only the status inference of the current parking space but also the inference of the adjacent parking space. Even though the classification result at a single image patch may not be always correct, we can combine the classification results of several image patches around a parking space to achieve more robust inference. Given a parking lot, we first set up an IP camera on the roof of a building near the parking lot. The camera is geometrically calibrated to obtain the 3-D to 2-D projection model and to construct the 3-D plane-based scene model. After that, we capture a few image sequences of the monitored parking lot and extract plentiful image patches for each type of planar surface. For each image patch, we manually collect its patch label and extract its HOG feature from the normalized patch. Based on the labeled patch type and the HOG feature, we learn the conditional probability function p(o|l), where o denotes the observed feature of an image patch and l ∈ L denotes the label of the image patch. Since the surface type of an image patch can always be determined based on the 3-D scene model and the 3-D to 2-D transformation, we simply construct the conditional probability model for each of the four surface types. Before classification, we apply the multi-class LDA over the training image patches of each surface type to reduce the high-dimensional HOG features down to a much lower dimension. Taking the learning process of the surface type T as an example, each image
References


Chang-Chun Huang (M’09) received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 2000, 2002, and 2010, respectively. He is currently an Assistant Professor with the Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Taiwan. His current research interests include image/video processing, computer vision, and computational photography.

Yin-Shiu Tai received the B.S. degree in engineerin-

Sheng-Jyh Wang (M’95) received the B.S. degree in electronics engineering from National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1999 and 1995, respectively. He is currently a Professor with the Department of Electronics Engineering, NCTU. His current research interests include image/video processing, computer vision, and image analysis.

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 23, NO. 9, SEPTEMBER 2013