Automatic Vacant Parking Places Management System Using Multicamera Vehicle Detection

Rafael Martín Nieto, Álvaro García-Martín, Alexander G. Hauptmann, and José M. Martínez

***Abstract*— This paper presents a multicamera system for vehicles detection and their corresponding mapping into the parking spots of a parking lot. Approaches from the state-of-the-art system, which work properly in controlled scenarios, have been validated using small amount of sequences and without more challenging realistic conditions (illumination changes and different weather conditions). On the other hand, most of them are not complete systems, but provide only parts of them, usually detectors. The proposed system has been designed for realistic scenarios considering different cases of occlusion, illumination changes, and different climatic conditions; a real scenario (the International Pittsburgh Airport parking lot) has been targeted with the condition that existing parking security cameras can be used, avoiding the deployment of new cameras or other sensors infrastructures. For design and validation, a new multicamera data set has been recorded. The system is based on existing object detectors (the results of two of them are shown) and different proposed postprocessing stages. The results clearly show that the proposed system works correctly in challenging scenarios including almost total occlusions, illumination changes, and different weather conditions.**

***Index Terms*— Parking management system, vehicle detection, homographies, perspective correction, automatic spot mapping, multicamera fusion.**

I. Introduction

PARKING lots are a widely used service where a great investment is made every year. The management of these car parks is very expensive and in many cases complex, especially in the case of those that have many places such as airports or large commercial areas. Solving this problem using computer vision promises a number of advantages over intrusive sensors like induction loops or other weight-in-motion sensors [1]. In addition, a vision-based system may provide many value-added services, like parking space guidance and video surveillance [2]. Such systems allow the decongestion of crowded parking areas, directing vehicles to areas with lower occupancy, guiding the vehicles by a faster route.

Manuscript received March 27, 2017; revised November 2, 2017, February 6, 2018, and April 3, 2018; accepted May 11, 2018. This work was supported in part by the Spanish Government FPU Grant Program (Ministerio de Educación, Cultura y Deporte) and by in part by the Spanish Government under Project TEC2014-53176-R (HAVideo). The Associate Editor for this paper was J. Zhang. *(Corresponding author: Rafael Martín Nieto.)*

R. Martín Nieto, Á. García-Martín, and J. M. Martínez are with Universidad Autónoma de Madrid, 28049 Madrid, Spain (e-mail: rafael.martinn@uam.es).

A. G. Hauptmann is with Carnegie Mellon University, Pittsburgh, PA 15213, USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2838128

Surveillance cameras are readily available in most car parking lots, so in many cases the solution is only to adequately process the information available from the already existing cameras, or complete the deployment by adding some cameras to have a full coverage that allows the system to operate.

The previously developed systems are mainly based on image segmentation or machine learning (SVMs, NN) over spot patches, but due to the evolution in the last years of object detection algorithms, it is possible to use the detections of these algorithms for the proper operation of automatic parking management systems.

This paper is structured as follows: after this introduction, section II presents an overview of the related work. Section III presents and describes details of the complete system and each of the blocks that compose it. Section IV presents the evaluation framework (dataset and evaluation metrics) used to obtain quantitative results of the system. Section V presents the experiments and results obtained by the system. Finally, section VI describes the conclusions of the paper and some lines of future work.

II. State of the Art

In this section, we overview works related to the proposed automatic parking management system, which try to locate occupied/empty parking spots. We have organized all the related works in three categories taking into account the technique used for the occupied/free parking spots classification: image segmentation, machine learning (SVMs, NN, etc.) over spot patch (or patches), and vehicle detection techniques based on object detectors.

A. Image Segmentation Based Systems

Image segmentation based systems try to differentiate, in each considered frame, between vehicles and parking spots. Background subtraction is a typical technique used in this category, where an empty image is used to subtract each frame in order to get the foreground mask (vehicles). The vehicles are extracted and then mapped to each parking spot. The most representative works included in this category are [1]–[13]. The algorithm from [1] considers three main processing stages: firstly, shadows in the image are attenuated (or removed) and image distortion is corrected; afterwards, correspondences are established between stationary cameras and visible parking places, and, finally, the parking place status is evaluated. Status classification is based on the assumption that the surface of a

vacant parking place is relatively invariant in comparison to an occupied place. The parking slots labeling process is treated in [2] as a color classification process which decomposes the image observation into an object component and a lighting component. The object type is either "car" or "ground", and the lighting condition is either "shadowed" or "unshadowed" (the system needs to know the direction of sunlight). Both the expected object map and the expected shadow map are created to help in the image pixels labeling. A frame preprocessing is applied using the Surface Texture and Microstructure Extraction (STME) in [3], resulting in an image where the vehicles appear as "bumps" in an elevation map. A method for individual vehicle detection using grayscale images acquired from an elevated camera is presented in [4]. Vehicles are considered to be composed of several components such as hood, windows, headlights, etc., so images of parking cells are fragmented by gray level and a cell is considered occupied if it is composed by a large number of small components. A dual camera device was designed and calibrated manually in [5], where parking lots (manually specified) are detected using background subtraction. After that, two morphological operations, erosion and dilation, are performed to connect the blobs and to eliminate the noise. The system introduced in [6], and enhanced in [7], needs to store an unedited zero occupancy image, and manually store the identified coordinates of every parking spot. The object (vehicle) detection is based on a combination of background extraction and edge detection (using the Sobel operator). Background subtraction is also used in [8] but with two additional considerations: a preprocessing color filter is applied for maintaining color stability, and a shadow removal is used to remove shadow foreground pixels. A spot is considered occupied if the percentage of the foreground pixels in the spot patch are over an empirical threshold. A stitching algorithm is used in [9] to integrate visual cues from multiple cameras for constructing a panoramic scene. Color, position and motion are used for tracking vehicles across different cameras. Two features are used to capture the vacant properties of each parking space: edge (Canny filter) and color (background subtraction). Three different methods of image analysis are combined in [10]. Edge counting and histogram classification are utilized as static analysis methods (information available in a single frame), and a crafted algorithm for blob tracking as dynamic (across-frames) method using background/foreground estimation. The vehicular occlusion problem, which is important in other approaches, is supposed to be avoided in the paper through camera placement at high floors, which is not always a possible solution. In [11], after an initial edge detection stage, edge density, closed contour density and foreground/background pixel ratio are combined to decide whether a car is present or not in each parking spot. The parking space boundaries are fixed, and the region of each parking space can be defined using 4 dots or just a parallelogram as a given parking spot. After that, each parking space is numbered. The parking management system described in [12] tries to find the car park coordinates from an empty car spot, acquiring an image with cars, converting the image to black and white for simple analysis, removing noise and determining whether car spots are vacant or filled. Each spot is segmented to decide whether it belongs to the background (empty) or to the foreground (occupied). Two types of car parking lots photos are used: one is taken from Google earth and the other one is a real car park photo. After a homography transformation, the system presented in [13] performs a background subtraction, and a feature classification (SURF [14] and HOG [15]) to decide the status of each parking spot.

B. Spots Patch Classification Based Systems

Spots patch classification based systems use classification machine learning techniques (SVM, NN, etc.) which are trained with previously labeled patches of occupied and free parking spots. The most representative works included in this category are [16]–[25]. The parking management system presented in [16] creates, using homography computation, a pseudo-top-view of a parking area to determine if there are free parking lots or not. The texture feature extraction of each parking lot is obtained using Gabor filter banks. A SVM is trained with texture feature vectors of every parking spot, which have been taken in different illumination conditions and with diverse type of shadows. The algorithm proposed in [17] uses a combination of car feature point detection and color histogram classification to detect vacant parking spots. The author points out that one major weakness of this algorithm is that it cannot accurately detect the state of parking spots which are slightly or mostly occluded by objects such as other vehicles. The method for parking space detection proposed in [18] trains and recognizes empty parking spaces by applying machine learning methods (SVM). Three consecutive parking spots are proposed as a detection patch, which contains the space under consideration and the two neighboring spaces. The system uses PCA to pick 50 critical features. The problem is addressed in [19] through a Bayesian hierarchical detection framework. The top layer is an observation layer, where each node indicates a local feature. The local feature can be either texture-based or pixel-based. Haar-like features are used in [20] for the detection of features detected in input videos to determine the presence of a car within a parking spot. A surface-based hierarchical framework is proposed in [21] to integrate the 3-D scene information with the patch-based image observation for the inference of vacant space. The HOG feature dimension is reduced using a Linear Discriminant Analysis (LDA), and 4 likelihood models are trained for each surface type. A classification of several algorithms for vacant parking space detection is presented in [22], depending on the challenges that they consider for vacant parking space detection: perspective distortion, inter-object occlusion, shadow effect, lighting variations and insufficient illumination at night. They use HOG features for car detections, and cars are decomposed into four types of planar surfaces. Since the perspective projection process is highly dependent to the camera setting, the patch classification models need to be re-trained for different camera settings. It takes two days to install the system. The first day is used for hardware setup, camera calibration and training data collection. The second day is used to label the training data and to learn models

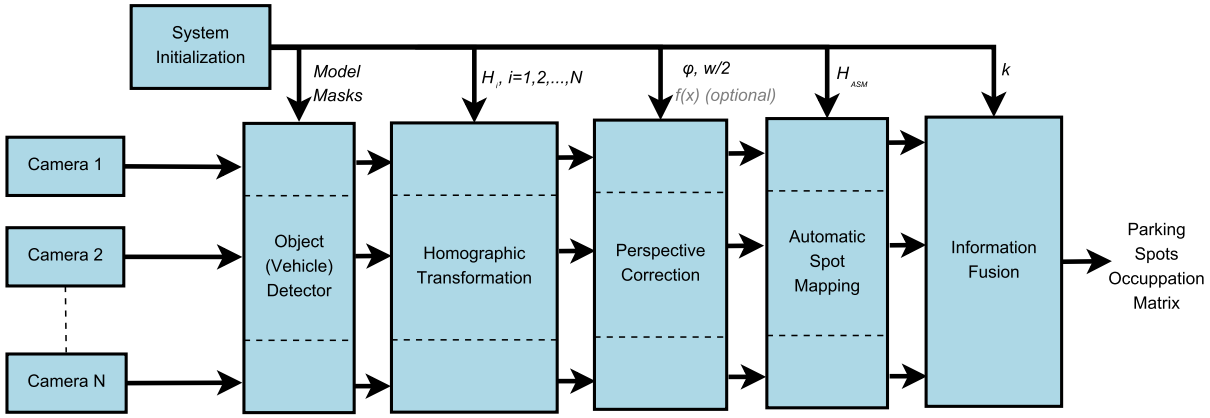

Fig. 1. System Block Diagram: cameras provide the frames to be processed (left); the initialization block provides the necessary information for each block (top); and the result of the system is the parking spots occupation matrix (right).

for patch classification. Most of the failure cases are caused by the headlight of moving cars. Reference [23] extends the system presented in [22] adding a multiclass boosting method to automatically select the weak classifiers weights through a back-propagation learning process. This system is divided into 3 layers: 3D-cuboid model and feature extraction layer, patch classifier layer, and weighted combination layer. Like in other systems already mentioned (e.g., [21]), a LDA process is used to reduce the feature dimension of the extracted HOG features. In [24], several features with different color histograms or DoG histograms are analyzed using three supervised learning algorithms (k-NN, LDA, SVM). Finally, a multi-layer discriminative framework for vacant parking space detection is presented in [25]. This extended framework adds a status inference layer over [22].

C. Object (Vehicle) Detectors Based Systems

Object (vehicle) detectors based systems use a detection algorithm to detect vehicles and to map them into the different parking spots. This type of system has begun to be viable in recent years thanks to the evolution of object detectors, specifically [26]–[28]. The only work, based upon our knowledge, included in this category is a car detection method [29] based on the Convolutional Neural Networks (CNN) technology. After training the CNN, to identify where there are cars they search the whole image of a parking lotusing a sliding window approach. In this work the detection is performed but the results are not mapped in the different parking spots, and therefore it is not a complete system.

In our work we also propose to follow the detection approach but designing and developing the different stages to get a complete automatic parking management system: vehicle detection, homographic transformation, perspective correction (for allowing to reuse existing camera installations), automatic spot mapping and multicamera fusion (assuming the usual availability of multicamera setups). Additionally we have created a complete realistic dataset including a multi-camera environment with both illumination and climate variability and we perform a rigorous and methodological evaluation of the proposed system.

D. Qualitative Comparison Between Existing Approaches and the Proposed System

Due to the absence of public datasets of stationary vehicles, it is not possible to make a quantitative comparison of the proposed detection based system with respect to the others, however, the novelty of the proposed detection based system allows to conceptually compare the advantages of the system compared to the existing ones.

An advantage of the proposed system over existing systems is the "automatic vehicle mapping" on the different parking spaces. Many approaches (e.g., [7], [8], [12], [16], [17], [20]) require manually annotating, one by one, the position of each spot, while our system needs only the corners of the parking area and the number of spots. This advantage is especially notable compared with the spots patch classification based systems and especially in the case of large car parking in which the number of places to label is high. The main advantage of the proposed system over the image segmentation based systems is the robustness against variable background, generally caused by climatic or lighting variability. This system is the first of its class to detect and subsequently map in the different parking spaces, as [29] just detect the vehicles and does not perform the subsequent steps.

Another advantage of detection based systems is the capacity to withstand "object occlusions". Although some of the existing systems (e.g., [18], [22], [23], [25]) already try to support occlusions, the object detectors have a better capacity to support them because they use the information they have without needing to add dependency occupation rules between adjacent spots.

Finally, adding "multicamera support" to the system allows the existence of complete occlusions in the scenario, and the use of redundant information from the different cameras allows to improve the system performance.

III. Proposed System

A. Overview

The proposed multicamera system is based on a parallel processing of each camera followed by the combination (or fusion) of their individual results. The block diagram of the system is presented in Figure 1. Each camera captures frames,

(a) (b) (c)

Fig. 2. Example of (a) ROI mask, (b) input frame and (c) masked frame.

which are processed frame-by-frame. Firstly, an "object detector" (using a previously trained vehicle model) locates the vehicles in the frame; using an "homographic conversion" and a "perspective correction" to consider the volumen of the detected objects, the obtained detections are "automatically mapped" into the positions of the occupied/empty monocamera spot matrix. Finally, if there is a mutlicamera setup, the information from each camera is "fused" to obtain the final multicamera spot matrix which indicates the occupation of the parking lot.

In order to present a system configuration example, we proceed to describe the source of each of the modules of our implementation: the considered detectors are existing techniques from the state of the art but their models have been trained specifically for the purposes of the system; the homographic transformations are mathematical techniques described in [30] but we have generated our own homography matrix for each of the cameras; the perspective correction is a technique designed by us for this system and is based on trigonometry; the automatic spots mapping is a technique designed by us based on homographies; the fusion considers tuned functions of standardized sigmoids designed for our interests.

B. Object (Vehicle) Detector

The object vehicle detector is initialized with the vehicle model, and, in order to eliminate possible detections of other areas of the parking lot that will not be monitored with these cameras, it also receives a region of interest (ROI) mask for each camera. An example of these masks is shown in Figure 2.

This block receives the frames of each camera and, using an object detection algorithm, generates as output a bounding box (rectangle) for each of the detected objects (vehicles).

We have trained new vehicle models due to existing car models do not function properly when using an image with a high viewpoint, scales variability, occlusions, different vehicle types, etc., as contemplated in the experiments sequences.

The main detection algorithm selected for the evaluation of the proposed system is the Faster R-CNN (Regions with Convolutional Neural Network Features) [28] detector, which is a more efficient variation, mainly in terms of computational cost but also in performance, of the previous R-CNN [26] and Fast R-CNN [27] detectors. The three variations have in common the combination of bottom-up region proposals with rich features computed by a convolutional neural network. The main difference of the Faster-RCNN is the use of a Region Proposal Network (RPN) that enables nearly cost-free region proposals. For training purposes and according to the author's results [28], we have chosen the pre-trained network VGG-16 model [31] that has 13 convolutional layers and 3 fully-connected layers. We have retrained the network using the PASCAL VOC 2007 and 2012 datasets and we have added a new object model, our parking vehicle model, using our dataset (see section IV-A). The vehicle detector used in [29] is also based on a generic CNN from the state of the art, trained by the authors. As the code and model are not publicly available, it cannot be used in the evaluation of the system, but it could be integrated and evaluated in a direct way

The second detection algorithm evaluation is the Deformable Parts Model (DPM) detector [32]. The DPM detector is based on exhaustive search and a part-based model. It is a part-based adaptation of the original Histogram of Oriented Gradients detector (HOG) [15]. It proposes an object detection system based on mixtures of multiscale deformable part models where each deformable object part is modeled as the original HOG detector [15]. The algorithm model also contains the flip of the model. We used this detector in order to see the behavior of the system when using a non deep-learning based detector. As deep-learning based ones are "better" detectors, this evaluation allows to demonstrate the robustness of the system to detection noise.

Additionally, experiments were also made with the ACF (Aggregate Channel Features) [33] algorithm, but, due to the properties of this technique, the bounding boxes obtained during the detection process covered only the roof of the vehicles, instead of completely covering them. This causes that this algorithm does not fulfill the requirements for the detectors of the proposed system which considers that the bounding boxes completely cover the vehicle.

C. Homographic Transformation

The object detector of the previous block obtains a bounding box for each detected vehicle from the viewpoint plane of each camera. This block, using the properties of the homographies, allows to change the position of the objects detected from the plane of each camera to a common plane. The homography matrix (which is needed to initialize the block), H_i, for each camera i, is obtained using 4 points from each camera viewpoint and each point correspondence in an image extracted from a top view. This top view can be easily obtained from *Google Earth*. It is not necessary to choose the same points in each camera viewpoint for all the cameras, but each selected point must be associated with one from the image of the common ground plane. The dimensions of the matrices are, by definition of homographies, 3x3. Figure 3 shows two examples of the resulting viewpoint change using homographies. Note that these images are generated only to illustrate the procedure, but this computationally expensive step is not required during the system operation by the mapping algorithm. Therefore, homography is just applied to the base midpoint of each bounding box resulting in an optimized computation. The output of this block is one point for each detected vehicle.

D. Perspective Correction

Due to the volume of the detected objects, it is necessary to correct the positions of the projected points where the detections, received from the previous stages, are mapped.

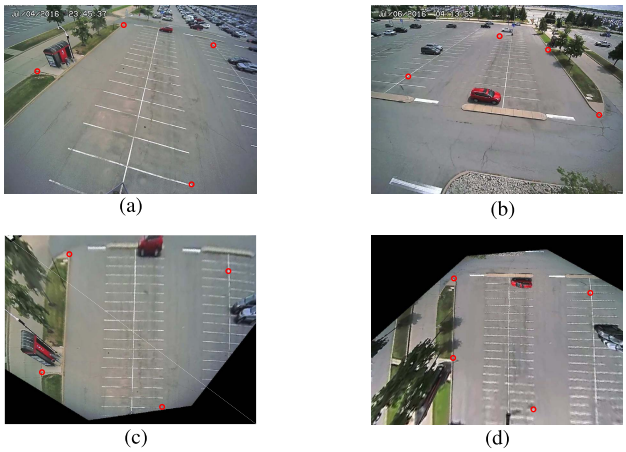

Fig. 3. Homography viewpoint transformations: (a) and (b) show the starting side viewpoints, (c) and (d) show the resulting top view common planes.

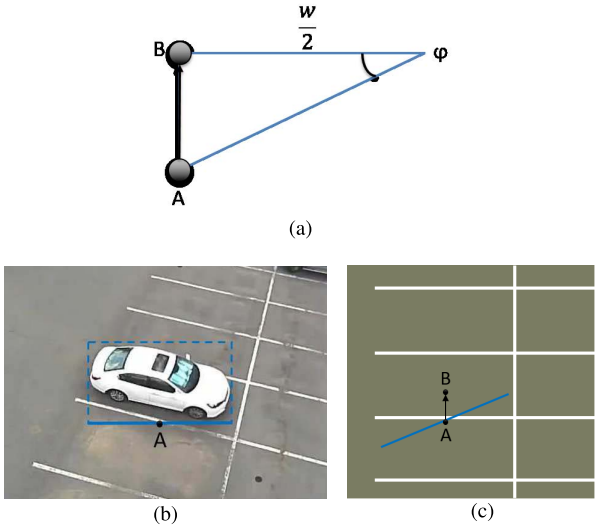

Fig. 4. Perspective correction diagram and example: (a) schematic diagram; (b) camera viewpoint detection; and (c) position correction example.

It is possible to make a position correction using the angle between the parallel lines of the parking spaces and the camera viewpoint. This allows the correct matching between the vehicle detections and the parking spots. Figure 4 presents the correction diagram and an example. In the diagram, A corresponds to the base midpoint detection projection, B corresponds to the final position after correction, φ is the angle between the parking lines and the camera view (needed for the block initialization), and $\frac{w}{2}$ is the half of a vehicle length (average). Despite referring to the length of the vehicle, the letter w is used to associate it with the width of the bounding boxes. In the example, the blue line in Figure 4(b) represents the base line projection of the detection bounding box. Note that the midpoint of the base is a different point than the center point of the bounding box; the midpoint of the base, belonging to the ground plane, allows to fulfill the properties of the applied homography. In addition, the choice of this point allows the system to work independently of the height at which the cameras are located, as its projection is always placed between the vehicle and the camera (see Figure 4(b)).

On the other hand, it is possible that the distortion of the lens affects the accuracy of the homography, which causes errors and imprecision in the mapping of the spots, as seen in the Figure 5(a). This problem is usually solved using the intrinsic camera parameters. If these parameters are not available, there is an alternative solution that consists of correcting the mapping of the grid of spots using a simple linear adjustment function (Figure 5(c) shows an example of adjustment function). Figure 5(b) shows the result of applying this correction. We define a uniform grid and then we add a correction factor to the projected point in order to eliminate the effect of the lens. You can correct the grid to fit the image, or correct the image to fit the grid. For the system, it is more efficient to correct the image to fit the grid, since it allows to automate the subsequent steps without needing any other correction. The grid could also be modified, but the image correction simplifies the next step of automatic mapping, as the uniform grid allows the automatic spot mapping via homographic techniques.

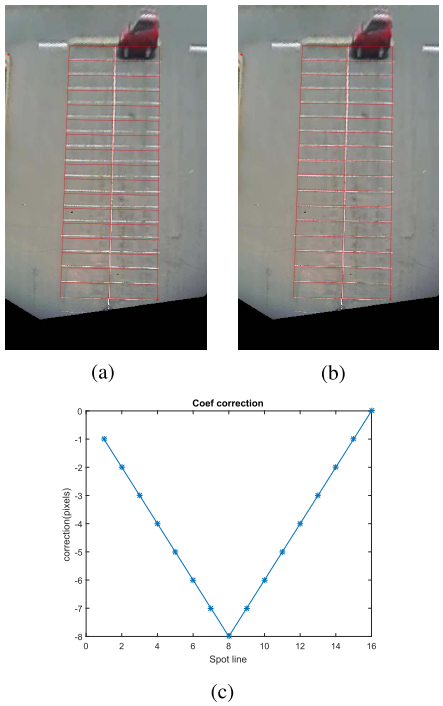

Fig. 5. Camera lens correction: (a) initial grid, (b) corrected grid and (c) correction function.

E. Automatic Spot Mapping

This block is based on using the properties of the homographies. However, in this case the selected destination points are designed specially to get the automatic discrete spot numbers directly without the need of supervision and without the need to map each position one by one like most of the state of the art systems (e.g. [7], [8], [12], [16], [17], [20]). The source points are the four corners of the parking grid, and the destination points are the corners of the synthetic destination space shown in Figure 6, specifically: $(1, 1)$, $(1, M + 1)$, $(N + 1, M + 1)$ and $(N + 1, 1)$. M is the number of parking columns (1 or 2),

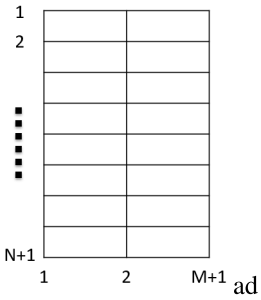

Fig. 6. Destination points for the automatic spot mapping.

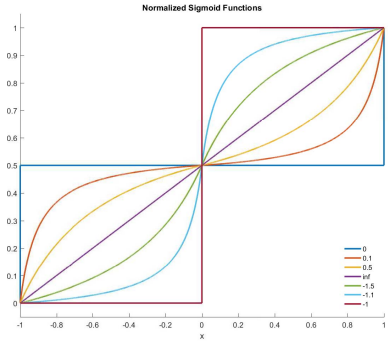

Fig. 7. Normalized sigmoid functions using different k parameter values.

and N is the number of parking rows. The procedure to obtain the discrete (d) position of the mapped detection in the occupation spot matrix (x_d, y_d) consists of two steps and is presented below:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = H_{ASM} \begin{bmatrix} x_{cp} \\ y_{cp} \\ 1 \end{bmatrix} \tag{1}$$

where, H_{ASM} is the homography matrix for automatic spot mapping, obtained with the previously defined source and destination points, x_{cp} and y_{cp} are the x and y coordinates of the projected (and corrected in the previous stage) detections mapped in the common plane. This matrix H_{ASM}, common to all cameras, should not be confused with the matrices H_i, defined for each camera and with different functionality (see Section III-C) This block needs H_{ASM} for its initialization.

Finally, the operation that allows obtaining the discrete value of the occupied spot is:

$$(x_d, y_d) = \left\lfloor \frac{x'}{z'}, \frac{y'}{z'} \right\rfloor \tag{2}$$

The outputs of this block are the occupancy spot matrices generated by each camera.

F. Information Fusion

Logically, due to the resolution of cameras, optics, etc., the greater the distance between the camera and the mapped detections, the lower the accuracy. In order to deal with this factor, it has been decided to study a method to fuse the information of all cameras using a normalized sigmoid function that allows to evaluate/study different combination approaches in a simple way (using a unique parameter). For this purpose, a normalized sigmoid function, $P(x)$, has been used:

$$P(x) = \frac{kx}{k - x + 1} \tag{3}$$

where, x is the normalized position between each camera and the center of the parking, and k is the parameter which allows to tune the sigmoid. The formula presented works for $0 < x \leq 1$, the normalized distance between the camera and the center point of the parking area. It is necessary to repeat the function for negative values, to get the range from -1 to $+1$. This is achieved by giving the function the absolute value of x, and then changing the sign of the result back to the same sign as x. Additionally the result is rescaled so that at the ends (x $= 1$ and x $= -1$) the function takes values of 1 or 0. The final normalized function, $P_{norm}(x)$, is defined as:

$$P_{norm} = \begin{cases} \dfrac{0.5kx}{k - x + 1} + 0.5 & 0 < x \leq 1 \\ 0.5 & x = 0 \\ \dfrac{-0.5k|x|}{k - |x| + 1} + 0.5 & -1 \leq x < 0 \end{cases} \tag{4}$$

Some resulting sigmoid functions are shown in Figure 7 with different examples for the k parameter. In this way, the camera whose detections are weighted is placed at point x $= 1$, and the center point of the monitored parking area is placed at point x $= 0$. In the case of systems with two cameras, the other camera is located at the point x $= -1$, but this weighting of the detection confidence does not require the system to use only two cameras since it supports any number of them. In a scenario with more than two cameras, it is necessary to define the center of all of them, and each camera will have an associated function $P_{norm}(x)$, adapted by its corresponding distance to the center.

As shown in Figure 7, it is possible to obtain the extreme cases in which a plane function ($k = 0$) is obtained with a constant value equal to 0.5 (all cameras have the same confidence for all the points of the parking), a step function ($k = -1$), and a straight line of slope 1 between $x = -1$ and $x = 1$ ($k = \infty$, only the nearest camera detections are considered). It is also possible to obtain symmetric curves with respect to the straight line of slope 1 (values 0.1 with -1.1, and 0.5 with -1.5).

Thanks to this confidence weighting, the most distant detections will lose score against those close to the camera. After this, the detections of all the cameras are added and are used to obtain the final parking spaces occupancy matrix. For the sigmoid with k $= 0$, the result is equivalent to adding all detections of all cameras with their original score, since all of them are weighted by a value of 0.5. In the case of k $= -1$, the only detections that are maintained in each camera are those of which the camera that detects is the closest of all of the cameras. This case is a combination of information between the different cameras, each covering the area which contains the nearest spots.

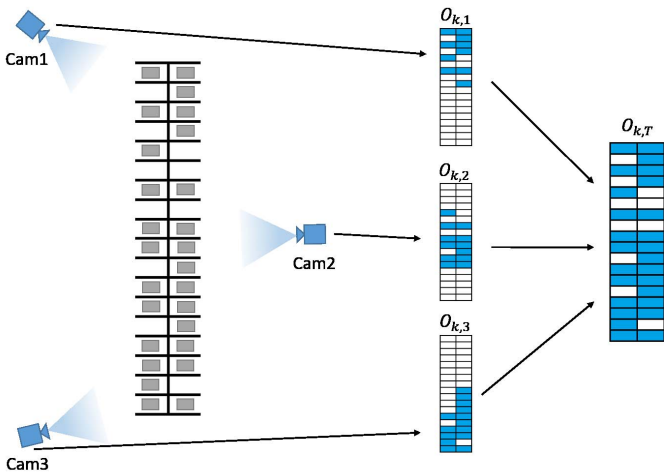

Fig. 8. Information fusion example: parking and monitoring cameras (left), monocamera extracted occupation information (center) and result of the information combination (right).

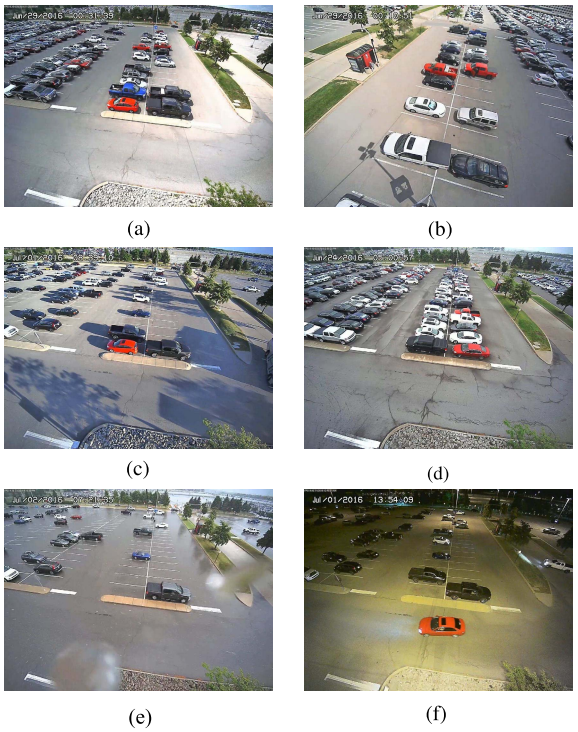

Fig. 9. Examples of dataset frames: (a) shows an example of Camera 1 viewpoint, (b) shows an example of Camera 2 viewpoint. (c)-(f) show examples of different illumination and weather.

For the selection of parameter k, the chosen detection algorithm and the scenario (mainly location of each camera) must be taken into account. Negative values should be considered for parameter k (e.g., −1, −1.1, -1.5) if the performance of the detection algorithm falls significantly with distance, or if the considered camera has low resolution, which complicates its detection. Otherwise, positive values of the parameter k (e.g., 1, 1.1, 1.5) will produce a better performance of the system as it considers the farther detections of each camera with greater weight.

After the automatic spot mapping (see Section III-E), the occupation matrix $O_{k,i}$ for camera i and a learned k parameter is defined as:

$$O_{k,i}(x_d, y_d) = \begin{cases} 1 & if\ spot\ (x_d, y_d)\ is\ occupied \\ 0 & otherwise \end{cases} \quad (5)$$

One occupation matrix is obtained for each camera. All of them are fused to obtain the total occupation matrix of the system, $O_{k,T}$:

$$O_{k,T} = \bigcup_{i=1}^{n_{cameras}} O_{k,i} \quad (6)$$

Figure 8 presents a simplified example of the complete process, in order to clarify the information fusion stage. In the left side, there is an example of parking with occupation, and three cameras monitoring the area of interest. In the center, the occupation information extracted by each camera is processed and the monocamera occupation matrix is generated. The right part of the example presents the result of the information combination, obtained by combining the information from all cameras.

IV. Experimental Setup

A. Parking Lot Dataset

The Parking Lot dataset (PLds) dataset was recorded as there was a lack of public parking lot datasets. We used it to test the designed system. The sequences were recorded in a real environment (Pittsburgh International Airport parking lot), in order to work with an environment as realistic as possible. Each frame, recorded using *Panasonic WV-SW155* cameras, has a resolution of 1280x960 pixels. Figure 9 shows an example of each one of the two viewpoints (9a and 9b), and examples of different illumination (day, night, sunrise with shadows) and weather (sunny, rainy) conditions.

The dataset consists of two main image sets, a training set used to generate the detector models, and a test set used for the experimental evaluation. The training set consists of a longer set of images, and the test set consist of a long (named All_CamX) and a short (named Synchronized_CamX) version of the images, with 1000 and 100 frames, respectively. The short versions (Synchronized sets) are subsets of the long versions: they consist of frames synchronized between the two cameras, to be able to evaluate the multicamera setup. The different image sets details are presented in Table I.

In addition to generating the images, the vehicles of all images have been manually annotated. The training images have been annotated for its use in the generation of the parked vehicle model, and the test images for the evaluation of the parking vacant management system. In the case of the Synchronized set, the vehicle occupancy matrix has been manually generated to also evaluate the system at this level. More details of this evaluation of the system are presented in subsection IV-B.

This dataset and its annotated ground truth are publicly available (http://www-vpu.eps.uam.es/DS/PLds/).

TABLE I
PROPERTIES OF EACH OF THE IMAGE SETS FROM THE PARKING LOT DATASET

Sequence name		#Frames	#Vehicles
Training		6616	28231
Test	All_Cam1	1000	12275
	All_Cam2	1000	9738
	Synchronized_Cam1	100	751
	Synchronized_Cam2	100	749

TABLE II
OCCUPATION MATRIX EVALUATION TABLE

Detected spot status	Ground Truth status	Spot evaluation
Vacant	Vacant	True Negative
Vacant	Occupied	False Negative
Occupied	Vacant	False Positive
Occupied	Occupied	True Positive

B. Evaluation Metrics

We quantify the performance results in order to evaluate the proposed approach. Global sequence performance is usually measured in terms of Precision-Recall (PR) curves [34]–[36]. These curves compare the similarities between the output and ground truth bounding boxes.

In addition, in order to evaluate not only the yes/no detection decision but also the precise object locations, we take into account the three evaluation criteria defined in [37], that allow to compare hypotheses at different scales: relative distance (dr), cover and overlap. A detection is considered true if $dr \leq 0.5$ (corresponding to a deviation up to 25% of the true object size) and cover and overlap are both above 50%.

The integrated Average Precision (AP) is generally used to summarize the algorithm performance in a single value, represented as the area under the PR curve (AUC-PR). In order to approximate the area correctly, we use the approximation described by [38].

In this paper two uses of the evaluation metric are distinguished. The first is the common one used for object detection, previously described. The second use is at occupied/empty spots level, according to the occupation matrix of the parking lot. Parking spaces may be occupied or empty. In this case it is a classification for each place, and the overlap is not measured for it. The occupation matrix and the Ground Truth are compared to define true positives, false positives, false negatives and true negatives, as shown in Table II.

V. EXPERIMENTS AND RESULTS

A. Detection Level Evaluation

As commented in section III-B, the first stage is performed by the vehicle detector.

We evaluate the detection results with and without the use of the ROI mask (see Figure 2) The detection results evaluation is done using the two detection algorithms presented in subsection III-B. All the generated models are executed over the four test image sets. Both detectors are evaluated with and without masking, e.g. Faster-RCNN default vs Faster-RCNN masked, in order to show its usefulness. The PR curves of this initial evaluation are shown in Figure 10 and the AUC values are shown in Table III.

All results of the masked detector are above those of the detections without masking. In particular, the Faster-RCNN detector is able to detect vehicles from other rows not controlled by the system and, therefore, not included in the manually annotated ground truth. This stands out for the camera 1, in which precision quickly falls as the recall increases.

If we compare the performance between the two cameras, in the case of the camera 1 the Recall performance usually decreases faster since it is positioned at a greater distance from the parking area controlled by the system. This will be taken into account in later stages by combining the information from the different cameras (see section V-C).

The All and Synchronized curves behave similarly, so the selection of synchronized frames is sufficiently representative for later experiments.

B. Perspective Correction Evaluation (Monocamera)

The second performed evaluation considers the perspective correction. This evaluation is carried out at parking spots level with the aim of demonstrating the need to correct the projection of the detections due to the point of view. Figure 11 shows the improvement of performing the perspective correction for both detectors, for each of the cameras, and including the corresponding precision-recall values obtained from the use of the Ground-Truth detection. Table IV presents the values of area under the curve (AUC) for all the curves shown in Figure 11. The results show that perspective correction is necessary and results in a significant improvement. From the Ground-Truth detection, the scores for each camera are also obtained (red and blue cross), which allow to have a measure of the best score that the monocamera detections can reach. Despite the improvement, the results can be further enhanced by the multi-camera information combination, which is presented in the following subsection.

C. Multicamera Information Fusion Level Evaluation

Finally, using the complete system, the matrix of occupied and empty spots is obtained and it is evaluated by comparing different results (the ones for each threshold of each detector model and the parameter k) with the ground truth of the matrix of occupancy of parking spaces. These results are shown in Figure 12. The Area Under the Curve of each detector is also computed and shown in Table V. This table also contains the result of the spots evaluation using the detections Ground Truth (manual annotations of bounding boxes for each camera view). Despite the use of ideal vehicle detections in this case, the result of the parking spots evaluation is not perfect (the value [1,1] is not reached for precision-recall) but the impact on the score is minimal ($<$0.03 precision lose). This error is due to the ideal annotations of vehicles contain subjective errors of the manual annotation (e.g., object bounding box estimation due to occlusions). We consider that it is not worth trying to correct it since it allows analyzing the impact on the result, which is despicable compared to the AUC

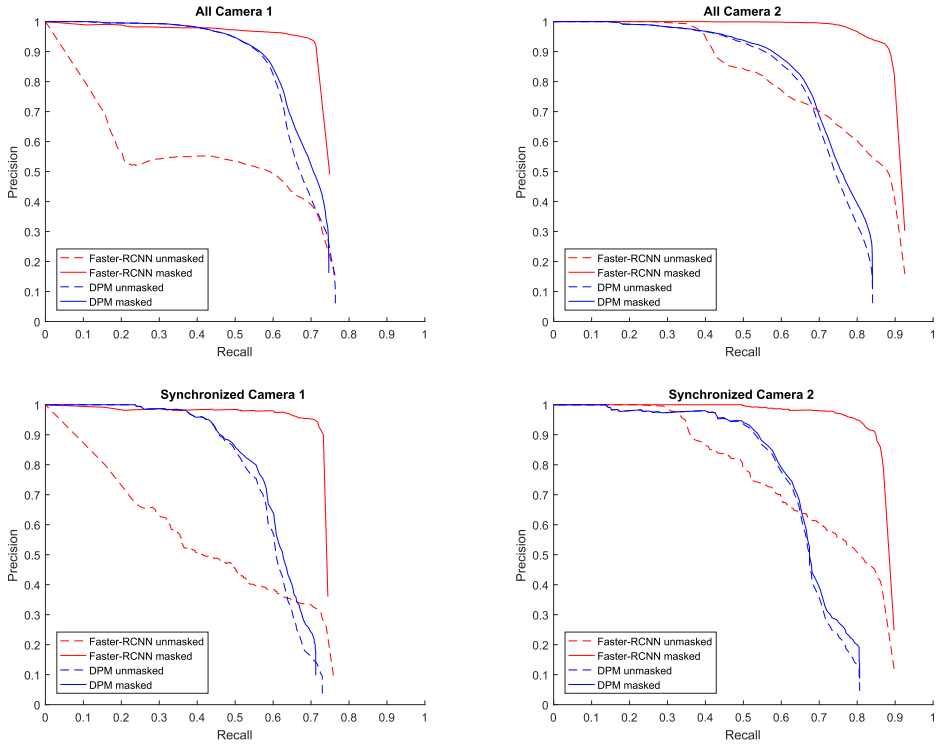

Fig. 10. Detection level evaluation for the two object detection trained models (Faster-RCNN [28] and DPM [32]).

TABLE III

AUC DETECTION SCORES FOR DETECTION LEVEL EVALUATION. THE BEST RESULTS OBTAINED FOR EACH IMAGE SET ARE SHOWN IN BOLD

Algorithm		All Cam1	All Cam2	Syn. Cam1	Syn. Cam2
Faster-RCNN	Unmasked	0.436	0.766	0.438	0.708
	Masked	**0.723**	**0.905**	**0.726**	**0.871**
DPM	Unmasked	0.664	0.717	0.598	0.661
	Masked	0.674	0.730	0.610	0.670

TABLE IV

MONOCAMERA AUC SCORES FOR PERSPECTIVE CORRECTION AT PARKING SPOTS LEVEL EVALUATION, FOR THE TWO OBJECT DETECTION TRAINED MODELS AND FOR THE TWO CAMERAS IMAGE SUBSETS. THE BEST RESULTS OBTAINED FOR EACH IMAGE SET ARE SHOWN IN BOLD

Algorithm		Syn. Cam1	Syn. Cam2
Faster-RCNN	Uncorrected perspective	0.380	0.452
	Corrected	0.526	**0.622**
DPM	Uncorrected perspective	0.440	0.402
	Corrected	**0.578**	0.537

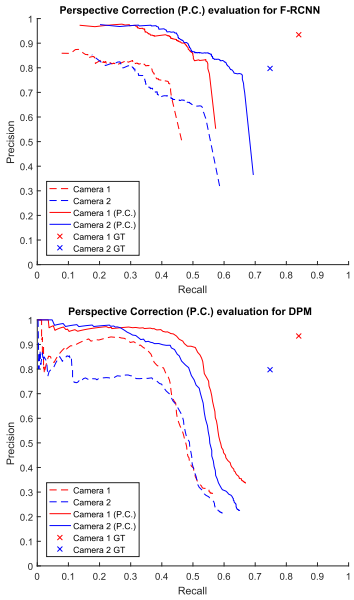

Fig. 11. Monocamera spots evaluation: perspective correction evaluation for the two object detection trained models and for the ideal detection (detection Ground Truth).

values obtained by the considered detectors. The best result is obtained with the Faster-RCNN detector, followed very closely by the DPM detector, but in all cases the results obtained are good enough for the proper functioning of the system (AUC around 0.9). Although the DPM detector presents worse results in detection (see section V-A), the complete system obtains, at spots level evaluation, very close results to those obtained by the Faster-RCNN detector. It is worth to point out the difference between the results of Figure V, which shows the spots evaluation at multicamera level, with the results of Figure 11, which shows the spots evaluation at monocamera level.

With respect to the different functions of normalized sigmoids used for the information combination/fusion, the results between them are very similar, but thanks to them it is possible to slightly improve the overall result of the system for a very small cost (simply weighting the detection scores of each bounding box depending on the distance to the detecting camera).

In order to configure a deployed system, the configuration of the parameters could be done in a convenient and simple way, as for the evaluation of spots it is necessary only to annotate a binary matrix of occupation (0 or 1 depending on whether the spot is empty or occupied). After this evaluation,

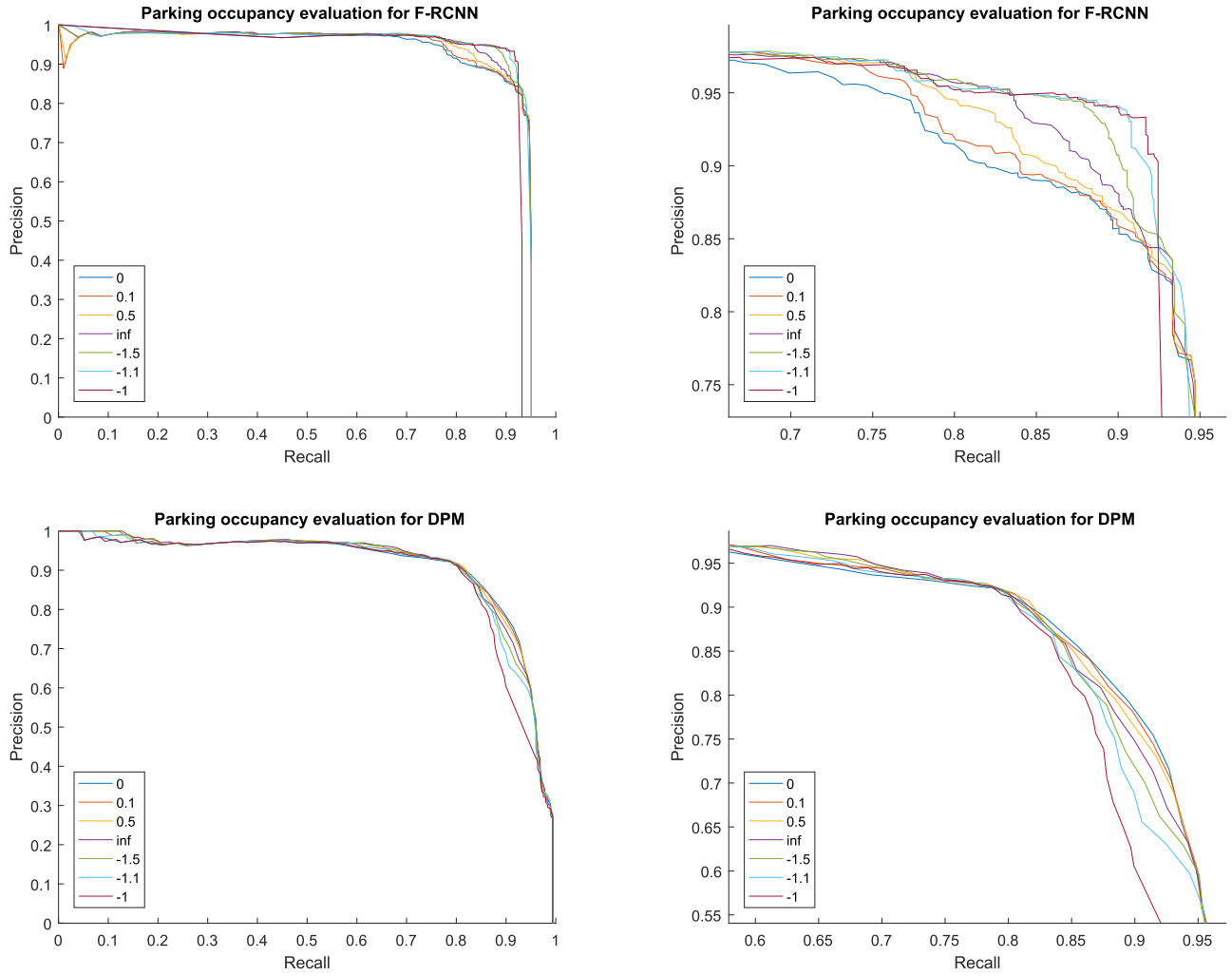

Fig. 12. Multicamera parking occupancy evaluation for the two object detection trained models: full curve (left) and zoom of the equal error rate area (right).

TABLE V

MULTICAMERA PARKING OCCUPANCY EVALUATION: AREA UNDER THE CURVE FOR THE TWO OBJECT DETECTION TRAINED MODELS AND FOR THE IDEAL DETECTION (DETECTION GROUND TRUTH), CONSIDERING DIFFERENT K PARAMETER FOR THE NORMALIZED SIGMOID FUNCTIONS. THE BEST RESULTS OBTAINED FOR EACH ALGORITHM ARE SHOWN IN BOLD

k	0	0.1	0.5	∞	-1.5	-1.1	-1	**Optimal k**
F-RCNN	0.910	0.909	0.912	0.916	0.918	**0.919**	0.905	-1.1
DPM	0.910	**0.913**	**0.913**	0.912	0.910	0.905	0.894	0.5
Det. GT	**0.991**	0.982	0.982	0.982	0.982	0.983	0.980	0

TABLE VI

PARKING OCCUPATION DENSITY EVALUATION: AREA UNDER THE CURVE FOR THE TWO OBJECT DETECTION TRAINED MODELS DIVIDED IN THREE OCCUPATION DENSITY CATEGORIES: LOW (1-12 VEHICLES), MEDIUM (13-24 VEHICLES) AND HIGH (25-36 VEHICLES)

		k						
Alg.	**Dens.**	0	0.1	0.5	∞	-1.5	-1.1	-1
F-RCNN	High	0.960	0.961	0.962	0.962	0.961	0.958	0.956
	Med.	**0.965**	**0.969**	**0.971**	**0.972**	**0.972**	**0.969**	**0.963**
	Low	0.863	0.867	0.870	0.871	0.872	0.869	0.860
DPM	High	0.918	0.919	0.919	0.920	0.920	0.919	0.892
	Med.	0.923	**0.924**	**0.924**	**0.925**	**0.926**	**0.928**	0.914
	Low	**0.930**	0.916	0.917	0.920	0.920	0.917	**0.926**

the weight of each camera would be learned from the k parameter and detection threshold with the best overall score. The computational and time cost of this evaluation is reduced and it does not require previous knowledge of the distance, reliability or quality of each camera for the person who is responsible for deploying and adapting the system.

1) Parking Occupation Density Evaluation: To verify that the proposed complete system is robust to occlusions, an additional study is added, classifying the frames in three occupancy density categories: low (1-12 vehicles), medium (13-24 vehicles) and high (25-36 vehicles). The results of this study are shown in Table VI. In spite of the need to divide a small number of frames (100) into three categories, which results in low resolution curves, the results show that the system performs correctly in occlusions situations. The results obtained by disaggregating the dataset are close to the mean except for Faster-RCNN low density occupancy, suffering a fall of performance of about 10% due to in scenarios with low vehicle density, a misclassification of a vehicle penalizes doubly (false positive and false negative) when occupying the square adjacent to the one it actually occupies. It should be noted that this system is especially designed for high density

TABLE VII

PARKING WEATHER EVALUATION: AREA UNDER THE CURVE FOR THE TWO OBJECT DETECTION TRAINED MODELS DIVIDED IN FOUR WEATHER CATEGORIES: DAYTIME/NIGHTTIME AND CLEAR/RAINY

		k						
Alg.	**Dens.**	0	0.1	0.5	∞	-1.5	-1.1	-1
F-RCNN	Day.	0.909	0.912	**0.913**	0.912	0.909	0.904	0.893
	Night.	**0.939**	0.935	0.927	0.924	0.930	0.931	0.928
	Clear	0.909	0.912	**0.913**	**0.913**	0.910	0.904	0.894
	Rainy	**0.917**	0.910	0.903	0.897	0.898	0.901	0.897
DPM	Day.	0.905	0.905	0.908	0.912	0.914	**0.915**	0.900
	Night.	**0.993**	**0.993**	**0.993**	**0.993**	**0.993**	**0.993**	0.972
	Clear	0.905	0.905	0.908	0.913	0.914	**0.915**	0.899
	Rainy	0.967	0.968	0.967	0.964	0.967	0.971	**0.973**

scenarios, where it is most useful to route vehicles to places where there are available spots.

2) Parking Weather Evaluation: Following the same procedure, a study of the system performance for different types of weather is added, classifying the frames in four weather categories: daytime/nighttime and clear/rainy. The results of this study are shown in Table VII. The scores obtained for the system in the nighttime frames are better than for the complete image set, due to during the night there are less reflections, which facilitates the operation of the detector. With respect to rainy frames, for the DPM detector the results get worse (between 0.003 and 0.015) for k = [0.1, 0.5, ∞, -1.5, -1.1] and improve (between 0.003 and 0.007) for k = [0, -1]. For the faster-RCNN detector, the results improve the overall performance between and 0.052 and 0.058 for all k. For daytime and clear frames sets, the behavior of the system is practically identical to the general behavior, as these categories contain most of the frames considered in the synchronized category. These small performance variations do not affect the system operation so, as indicated above, the system works for different types of lighting and weather conditions.

3) Optimal Parameter k: The process of learning the optimal k parameter for each algorithm consists of evaluating the range of values of the parameter, selecting the value that best adapts to the characteristics of the detection algorithm. After performing the experiments, considering the different possible values of k parameter, an optimal parameter has been obtained for each of the algorithms considered, as indicated in Table V. For the Faster-RCNN algorithm, the best sigmoid is obtained with the parameter $k = -1.1$. This is due to the most useful information is generated in the spots closest to each camera, and for this reason this parameter generates the best score after the dataset evaluation. In the case of the DPM algorithm, the best sigmoid is obtained with the parameter $k = 0.5$. In this case, the combination whose experimental score is better is obtained by maintaining the detections of medium distance with a greater weight than that considered for the Faster-RCNN algorithm. The optimal examples of sigmoids, and other examples, are shown in Figure 7. An optimum value is also obtained experimentally with k = 0 for the case in which the detections were ideal. This result is consistent as in the case of ideal detections, all detections have the same confidence and are, therefore, weighted with a constant value (flat sigmoid).

VI. CONCLUSION

This paper presents a multicamera system for the management of vacant parking places by means of vehicle detection and their corresponding mapping into the parking spaces of a parking lot. The system has been designed so that existing parking lot security cameras can be used for the proposed system after a simple configuration, without the need for a complete new camera deployment. The designed system faces more complicated scenarios than the ones tackled in the state of the art: almost total occlusions and climatic changes (cloudy scenarios, rain, snow...), that limits/reduces their performance. In this scenario with such a variable background it is not possible to carry out a precise background extraction, nor it is possible to label and define the region of each place as some parked vehicles completely occlude some of the spots behind them. In addition, the consideration of a multicamera scenario, which, as far as we know, has not been reported before for this type of systems, is added.

A new dataset has been recorded and synchronized. The publicly available dataset is composed by the generated parking vehicle models, the recorded frames and the ground truth files.

There are multiple future work lines to improve the proposed system. With respect to the combination, we have chosen a simple technique using normalized sigmoid functions, therefore different functions could be studied in order to optimize the combination or fusion of the different information sources. Also a new dataset with more cameras and with different spatial configurations could be recorded to see the behavior of the system in those situations. A tracker can be added to the sequence detection to combine the information extracted during the sequence frames providing temporary continuity to the vehicle detections. Apart from this, current lines of future work for object detection can be applied here, since the detector is the first stage of the system.

ACKNOWLEDGMENT

This work was developed during a stay at Carnegie Mellon University, thanks to the supplementary aid for FPU beneficiaries: Short stays.

REFERENCES

[1] T. Fabian, "An algorithm for parking lot occupation detection," in *Proc. Comput. Inf. Syst. Ind. Manage. Appl.*, Jun. 2008, pp. 165–170.
[2] C.-C. Huang and S.-J. Wang, "A hierarchical Bayesian generation framework for vacant parking space detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1770–1785, Dec. 2010.
[3] X. Wang and A. R. Hanson, "Parking lot analysis and visualization from aerial images," in *Proc. 4th IEEE Workshop Appl. Comput. Vis. (WACV)*, Oct. 1998, pp. 36–41.
[4] K. Yamada and M. Mizuno, "A vehicle parking detection method using image segmentation," *Electron. Commun. Jpn.*, vol. 84, no. 10, pp. 25–34, Oct. 2001.
[5] C.-H. Lee, M.-G. Wen, C.-C. Han, and D.-C. Kou, "An automatic monitoring approach for unsupervised parking lots in outdoors," in *Proc. 39th Annu. Int. Carnahan Conf. Secur. Technol. (CCST)*, Oct. 2005, pp. 271–274.
[6] D. B. Bong, K. C. Ting, and N. Rajaee, "Car-park occupancy information system," *Real-Time Technol. Appl. Symp.*, Apr. 2006, pp. 1–4.
[7] B. L. Bong, K. C. Ting, and K. C. Lai, "Integrated approach in the design of car park occupancy information system (coins)," *Int. J. Comput. Sci.*, vol. 35, no. 1, pp. 1–8, Feb. 2008.
[8] S. F. Lin, Y. Y. Chen, and S. C. Liu, "A vision-based parking lot management system," in *Proc. Int. Conf. Syst., Man Cybern.*, vol. 4, Oct. 2006, pp. 2897–2902.

[9] 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 *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2010, pp. 631–634.
[10] K. Blumer, H. R. Halaseh, M. U. Ahsan, H. Dong, and N. Mavridis, *Cost-Effective Single-Camera Multi-Car Parking Monitoring and Vacancy Detection towards Real-World Parking Statistics and Real-Time Reporting*. Berlin, Germany: Springer, 2012, pp. 506–515.
[11] L. Junzhao, M. Mohandes, and M. Deriche, "A multi-classifier image based vacant parking detection system," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst. (ICESC)*, Abu Dhabi, UAE, Dec. 2013, pp. 933–936.
[12] I. A.-B. Hilal Al-Kharusi, "Intelligent parking management system based on image processing," *World J. Eng. Technol.*, vol. 2, no. 2, pp. 55–67, Apr. 2014.
[13] I. Masmoudi, A. Wali, A. Jamoussi, and A. M. Alimi, "Vision based system for vacant parking lot detection: VPLD," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, vol. 2, Jan. 2014, pp. 526–533.
[14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. Comput. Vision Pattern Recognit.*, 2005, pp. 886–893.
[16] R. J. López-Sastre, P. G. Jiménez, F. J. Acevedo, and S. M. Bascón, "Computer algebra algorithms applied to computer vision in a parking management system," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jun. 2007, pp. 1675–1680.
[17] N. True, "Vacant parking space detection in static images," *Projects Vis., Learn.*, pp. 1–5, 2007.
[18] Q. Wu, C. Huang, S.-Y. Wang, W.-C. Chiu, and T. Chen, "Robust parking space detection considering inter-space correlation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 659–662.
[19] C.-C. Huang, S.-J. Wang, Y.-J. Change, and T. Chen, "A Bayesian hierarchical detection framework for parking space detection," in *Proc. Int. Conf. Multimedia Expo*, Mar. 2008, pp. 2097–2100.
[20] H. R. H. Al-Absi, J. D. D. Devaraj, P. Sebastian, and Y. V. Voon, "Vision-based automated parking system," in *Proc. Int. Conf. Inf. Sci., Signal Process. Their Appl.*, May 2010, pp. 757–760.
[21] C. C. Huang, Y.-S. Dai, and S. J. Wang, "A surface-based vacant space detection for an intelligent parking lot," in *Proc. 12th Int. Conf. ITS Telecommun. (ITST)*, Nov. 2012, pp. 284–288.
[22] C.-C. Huang, Y.-S. Tai, and S.-J. Wang, "Vacant parking space detection based on plane-based Bayesian hierarchical framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1598–1610, Sep. 2013.
[23] C. C. Huang, H. T. Vu, and Y. R. Chen, "A multiclass boosting approach for integrating weak classifiers in parking space detection," in *Proc. Int. Conf. Consum. Electron.*, Jun. 2015, pp. 314–315.
[24] M. Tschentscher, C. Koch, M. König, J. Salmen, and M. Schlipsing, "Scalable real-time parking lot classification: An evaluation of image features and supervised learning algorithms," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2015, pp. 1–8.
[25] C.-C. Huang and H. T. Vu, "A multi-layer discriminative framework for parking space detection," in *Proc. IEEE 25th Int. Workshop Mach. Learn. Signal Process.*, Sep. 2015, pp. 1–6.
[26] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. Conf. Comput. Vision Pattern Recognit.*, 2013, pp. 580–587. [Online]. Available: http://arxiv.org/abs/1311.2524
[27] R. B. Girshick, "Fast R-CNN," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448. [Online]. Available: http://arxiv.org/abs/1504.08083
[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. CVPR*, Sep. 2015, pp. 91–99.
[29] H. Xie, Q. Wu, B. Chen, Y. Chen, and S. Hong, "Vehicle detection in open parks using a convolutional neural network," in *Proc. IEEE Int. Conf. Intell. Syst. Des. Eng.* , Aug. 2015, pp. 927–930.
[30] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2003.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, Sep. 2014, pp. 1–14.
[32] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[33] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.
[34] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[35] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *Int. J. Comput. Vis.*, vol. 77, no. 1, pp. 259–289, May 2008.
[36] C. Wojek, S. Walk, and B. Schiele, "Multi-cue onboard pedestrian detection," in *Proc. IEEE CVPR*, Jun. 2009, pp. 794–801.
[37] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Jun. 2005, pp. 878–885.
[38] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2006, pp. 233–240.

Rafael Martín Nieto received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación degree) and the M.Phil. degree in research and innovation in information and communication technologies from Universidad Autónoma de Madrid, Spain, in 2012 and 2013, respectively. In 2007, he joined the Video Processing and Understanding Laboratory, Universidad Autónoma de Madrid. His research interests are focused in the analysis of video sequences for video surveillance (object tracking, object detection, multicamera systems, and so on).

Álvaro García-Martín received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación degree), the M.Phil. degree in electrical engineering and computer science, and the Ph.D. degree in computer science from Universidad Autónoma de Madrid, Spain, in 2007, 2009, and 2013, respectively. From 2006 to 2014, he was with the Video Processing and Understanding Laboratory, Universidad Autónoma of Madrid, as a Researcher and a Teaching Assistant. His research interests are focused in object extraction, object tracking and recognition, and event detection.

Alexander G. Hauptmann received the B.A. and M.A. degrees in psychology from The Johns Hopkins University in 1978 and 1982, respectively, and the Ph.D. degree in computer science from Carnegie Mellon University in 1991. For two years, he studied computer science at Technische Universität Berlin. In 1984, he was a Researcher with the CMU Speech Group, Carnegie Mellon University. His research interests are in speech recognition, speech synthesis, speech interfaces, and language in general. In 2003, he received the Allen Newell Award for Research Excellence, for the Informedia Digital Library, with H. Wactlar, M. Christel, T. Kanade, and S. Stevens.

José M. Martínez was born in Madrid, Spain, in 1967. He received the Ingeniero de Telecomunicación degree (six years engineering program) and the Doctor Ingeniero de Telecomunicación degree (Ph.D.) in communications from E.T.S. Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, in 1991and 1998, respectively. His professional interests cover different aspects of multimedia information systems, focusing on content analysis, understanding and description, and video summarization. He was a recipient of the Retevisión Award of the Official Professional College for Telecommunication Engineers to the best Ph.D. thesis of the academic year from 1997 to 1998.