Molina-Moreno, Miguel, González-Díaz, Iván, Díaz-de-María, Fernando. (2019). Efficient Scale-Adaptive License Plate Detection System. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), pp.: 2109-2121.

DOI: https://doi.org/10.1109/TITS.2018.2859035

Efficient Scale-Adaptive License Plate Detection System

Miguel Molina-Moreno, Iván González-Díaz, *Member, IEEE,* and Fernando Díaz-de-María, *Member, IEEE*

***Abstract*—License Plate Detection is a common problem in traffic surveillance applications. Although some solutions have been proposed in the literature, their success is usually restricted to very specific scenarios, with their performance dropping in more demanding conditions. One of the main challenges to be addressed for this kind of systems is the varying scale of the license plates, which depends on the distance between the vehicles and the camera. Traditionally, systems have handled this issue by sequentially running single-scale detectors over a pyramid of images. This approach, although simplifies the training process, requires as many evaluations as considered scales, which leads to running times that grow linearly with the number of scales considered.**

In this paper we propose a scale-adaptive deformable part-based model which, based on a well-known boosting algorithm, automatically models scale during the training phase by selecting the most prominent features at each scale, and notably reduces the test detection time by avoiding the evaluation at different scales. In addition, our method incorporates a empirically constrained-deformation model that adapts to different levels of deformation shown by distinct local features within license plates. As shown in the experimental section, the proposed detector is robust and scale- and perspective-independent, and can work in quite diverse scenarios. Experiments on two datasets show that the proposed method achieves a significantly better performance in comparison with other methods of the state-of-the-art.

***Index Terms*—License Plate Detection, GentleBoost, scale-adaptive part-based model, video surveillance.**

I. Introduction

With the increasing number of vehicles in urban and interurban roads, their identification becomes a key problem for several applications as security control (e.g. detecting stolen vehicles or traffic violations), traffic management, or organization of parking spaces. Automatic License Plate Recognition (ALPR) for vehicle identification is an essential module in many of these applications and has been a very active subject of research during the last decade [1]–[5]. The first stage in ALPR, known as License Plate Detection (LPD), entails the localization and cropping of the license plate area and has an important impact on the overall system performance.

License Plate Detection becomes challenging in unconstrained scenarios where no prior information can be used to drive the detection process. In such scenarios, many background objects might confuse the detector, the lighting conditions are strongly variable (24-hour operational video cameras), there is a large intra-class variation (e.g., American license plates have different designs for the different states), etc. Moreover, depending on the application, the size of the vehicles and, consequently, the license plate resolution, significantly vary according to vehicles distance to the camera. Thus, if not properly addressed, the scale problem may severely degrade the performance of the detector.

Diverse LPD systems are available in the literature [6]–[9]. Despite being efficient for constrained scenarios, these systems are significantly affected by varying illumination conditions (low-contrast license plates are hardly legible) or high-textured backgrounds. Some systems [10]–[12], although robust and efficient for low-quality images, are unable to handle even simple backgrounds showing other elements in the scene. Regarding the scale problem, morphology-based methods [6], [13], [14] rely on geometrical parameters, such as width, height or aspect ratio, to detect candidate regions to be license plates, which prevents an straightforward adaptation to a multi-scale problem. To address the scale problem, boosting-based methods (based on features, typically Haar, computed over an small window) sequentially run the detector over a pyramidal representation of the image [15], or scale up and down the feature window [16], [17]. In both cases, the computational complexity increases linearly with the number of evaluated scales.

In this paper, we propose a robust LPD system for unconstrained scenarios including day and night images and significant variations in viewpoint and scale. The system is built on the general object detector developed by Torralba et al. in [15], which uses the GentleBoost algorithm over a set of normalized correlation-based features. In comparison with this baseline detector, the main goal of our approach is to efficiently address the scale problem in multi-scale LPD systems. To this end, two are the main contributions in this paper:

- Instead of running the detector over a multi-scale pyramid representation of the image, we propose to divide the scale space into a discrete partition during the learning phase, and concurrently learn the best feature representation over all the scales. This allows us to take advantage of the correlation between adjacent scales and, during test, to perform the detection process just once over the image at its original resolution, thus avoiding the overhead due to the multi-scale processing and notably reducing the computational cost.
- We propose an empirically constrained-deformation part-

Manuscript received ... This work has been partially supported by the National Grants TEC2014-53390-P and TEC2017-84395-P of the Spanish Ministry of Economy and Competitiveness and SPIP2014-1507 of the Spanish Department of Traffic, Spanish Ministry of Home Affairs.

M. Molina-Moreno, I. González-Díaz and F. Díaz-de-María are with the Department of Signal Theory and Communications, Universidad Carlos III, Madrid, 28911 Spain e-mail: mmolina,igonzalez,fdiaz@tsc.uc3m.es.

based model to adjust the uncertainty regarding the relative location of the local parts within the whole license plate. The parameters of this model are learned on the training images and then incorporated to the detector, which makes it very robust and able to adapt to different scenarios.

The rest of the paper is organized as follows: in Section II we briefly summarize the related work. In Section III we describe in detail the proposed detector. In Section IV we compare its performance with that of some state-of-the-art approaches in two different tasks. Finally, in Section V we provide some conclusions and outline future lines of research.

II. Related work

LPD systems can be broadly organized into three groups: 1) those mainly based on morphological operations and character detection; 2) those based on boosting algorithms working on local features representing parts of the license, such as the one proposed in this paper; and 3) context-aware methods, which merge any of the previous alternatives with the detection of other parts of the vehicles, such as braking lights.

Regarding the first group, in [18] a detector relying on MSER (Maximally Stable Extremal Regions) [19] was proposed that achieves good performance, but it is very sensitive to lighting conditions (in particular, the detection rate varies between 83.3% and 95.6%). In [7], Chang et al. suggested a colour-edge-based detector, which focuses on certain edges where the colours match those found in license plates. A more recent approach by Fomani et al., described in [20], relied on a local histogram equalization and simple and adaptive morphological operations. In [6], [13], an initial candidate selection is performed by identifying areas with high density of vertical edges, which are then filtered using rules and constraints on basic geometric properties (aspect ratio, width or height of the candidate boxes). The main problems of these approaches is that they are not robust enough to varying lighting conditions and that they are not able to handle the scale problem, since many of the basic geometric parameters depends on the scale.

With respect to the second group, in [16] local and global features are combined in an efficient cascade detector. The two first stages of the cascade are based on global features (density and variance of gradients) and can discard up to 70% of the background of images, while the four following stages are based on local Haar features [21]. The reported detection rate in a multi-scale scenario (scaling up the feature window) is 93.5%; however, since it focuses on vertical edges it becomes quite sensitive to varying lighting conditions. Dlagnekov et al. [8] relied on the AdaBoost algorithm to select the 100 most significant features out of an initial set of 2400 Haar features. In this case, although the reported detection rate reaches a remarkable 95.6%, it should be noted that it addressed a simple task where the size of the plate remained fixed. A very recent approach [22] combines a pre-processing step based on morphological operations with a multi-layer hierarchical classifier with multi-scale block LBP (Local Binary Pattern) features, but again, it deals with a simple scenario where the images are taken from a fixed distance and the same point of view.

With respect to the third group, most of the systems rely on detecting some specific parts of the vehicle (such as headlights or braking lights) between which the license plate is located. In [23] a wavelet transform was used to locate the braking lights in the rear part of cars. After that, a more accurate method using mathematical morphology and geometrical clues was employed. Another approach [24] suggested using a part-based model with HOG (Histogram of Oriented Gradients) features to detect the vehicles, and a cascade classifier with HOG and LBP features to further detect the license plate. The relative license plate location along with other features were then used to classify the vehicles according to their make and model. However, these approaches either do not address the scale problem or simply use fixed constraints concerning license plates sizes or positions over the image (license plates are located in the bottom part of the image or on previously detected cars).

Recently, Convolutional Neural Networks (CNNs), which are currently the mainstream approach for tasks such as object detection or semantic segmentation, have already been applied to LPD [25], [26]. Li et al. [27] trained a 37-class CNN to detect all the characters in an image and a second CNN was used to process the resulting saliency maps. They reported notable results in LPD, with both recall and precision between 95 and 99%, depending on the task, but the system cannot be used in real-time. In [28] the problem is addressed by sequentially applying multiple CNN-detectors (car, plate and digits). The input images are processed with a Sobel operator to detect edges and a sliding window approach is used to propose regions to the first CNN, which is in charge of detecting cars. The second CNN acts on a sliding window over the regions provided by the first one to detect the license plates, and the third one works as a digit recognizer. Polishetty et al. [29] proposed a system which starts from Canny edge images. The Region of Interest (ROI) where the license plate is located is extracted relying on the edge images and those regions acts as inputs for a 9-layer CNN which classifies between license-plate/non-license-plate instances. Its reported performance (in terms of both precision and recall) varies between 91% and 99%. Finally, a recent approach [30] proposed a system based on faster R-CNN (Region-CNN) [31] where vehicles are detected first. After applying a pre-filtering based on relative size and aspect ratio of license plates, some candidates regions are proposed by means of Selective Search [32] to a second CNN. Generally, these systems are quite efficient in scenarios with few background objects.

The proposed approach differs from those described above in multiple aspects: first, the normalized correlation-based features used for the detection are robust against varying lighting conditions, which leads to a better performance in demanding environments. Furthermore, the discretization of the scale space allows us to limit the computational cost of the system at test time while still performing a multi-scale detection. Finally, the empirically constrained-deformation part-based model provides an improved model of the structure of the license plate which enables to address rich scenarios with complex backgrounds.

III. Proposed approach

The proposed detector is built on the general one by Torralba et al. [15], which relies on a deformable part-based model over local features and a boosting algorithm. As previously mentioned, we propose two innovations to the baseline model. First, we have designed a scale-adaptive part-based model which, through a discretization of the scale space avoids the search at several scales at test time, and takes advantage of the inter-scale correlations; and second, the relative position of each part of the part-based model with respect to the center is represented by means of a two-dimensional Gaussian distribution that allows us to properly model small spatial deformations according to training data.

Before describing our model in detail, we will introduce the baseline model [15] and discuss some particularities that arise from the application of such a general object detector to our particular LPD problem.

A. Baseline detector

The training process of the baseline detector [15] involves two steps: feature computation and learning of the boosting classifier (by means of the GentleBoost algorithm [33]).

1) Features: in order to extract the features, the first step is to build a vocabulary of visual words. For that end, a representative subset of training images is selected, and D patches $\mathbf{P}_f$, with $f = 1, \ldots, D$ are randomly extracted representing parts of the license plates. Let us note that, in order to improve the system performance and get invariance against variations in color and illumination, patches are extracted not only from the original image, but also from filtered versions. Hence, each patch is defined by the pair $(\mathbf{g}_f, \mathbf{P}_f)$, where $\mathbf{g}_f$ represents the filter applied to the image previously to the extraction of the patch. Next, the location $\mathbf{l}_f$ of each patch with respect to the center of the license plate is modeled by a two-dimensional Gaussian distribution (location mask), which allows some degree of deformation. Hence, a visual word in the vocabulary is finally defined by the triplet $(\mathbf{g}_f, \mathbf{P}_f, \mathbf{l}_f)$, including filter, patch and location. Figure 1 shows an example of a visual word.

Once the vocabulary has been built, we can compute the input features $\mathbf{v}^f$: they are similarity measures obtained by computing a normalized correlation between the filtered versions of the image $\mathbf{I}$ and the patches $\mathbf{P}_f$. A convolution with the transpose of the location mask $\mathbf{l}_f$ is then performed in order to refer the part scores to the center of the license plate. The process can be mathematically described as follows:

$$\mathbf{v}^f = (\mathbf{I} * \mathbf{g}_f) \circledast \mathbf{P}_f * \mathbf{l}_f^T \tag{1}$$

where $*$ and $\circledast$ stand for the convolution and normalized correlation, respectively. Figure 2 illustrates the whole process of training.

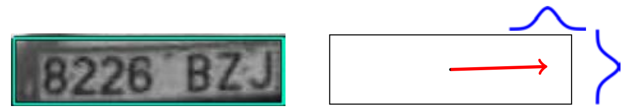


Fig. 1. Visual word illustration. Filtered patch (with an edge-detection filter) with their relative location respect to the center of the license plate (red arrow) and a degree of uncertainty (Gaussian distributions).

2) Boosting: the input features described above are used to train the detector implemented by means of a boosting algorithm. A boosting algorithm provides a way to sequentially add weak classifiers $\mathbf{h}_m$ in order to generate a strong classifier $\mathbf{H}$:

$$\mathbf{H}(\mathbf{v}) = \sum_{m=1}^{M} \mathbf{h}_m(\mathbf{v}), \tag{2}$$

where $\mathbf{H}(\mathbf{v})$ is the strong-learner for the feature vector $\mathbf{v}$, whereas $\mathbf{h}_m(\mathbf{v})$ is a weak-learner. The objective is to optimize the following cost function one term (of the additive model) at a time:

$$J = E\left[e^{-\mathbf{z}\mathbf{H}(\mathbf{v})}\right] \tag{3}$$

where $\mathbf{z}\mathbf{H}(\mathbf{v})$ represents the "margin", related to the generalization error; and $\mathbf{z}$ is the membership-label vector for each sample (± 1, for positive and negative samples, respectively).

In the literature, one can find several approaches to optimize this function, which leads to different versions of boosting: AdaBoost [34], GentleBoost [33], LogitBoost [33], etc. In particular, GentleBoost algorithm has been proved to be the most appropriate for license plate detection [35]. This algorithm performs the optimization of J following adaptive Newton steps, minimizing the square error in each step:

$$\arg\min_{\mathbf{h}_m} J(\mathbf{H} + \mathbf{h}_m) = \arg\min_{\mathbf{h}_m} E\left[e^{-\mathbf{z}\mathbf{H}(\mathbf{v})}(\mathbf{z} - \mathbf{h}_m)^2\right]. \tag{4}$$

If the expectation is replaced with an empirical average over the training data and the weights $w_i = e^{-\mathbf{z}_i \mathbf{H}(\mathbf{v}_i)}$ are defined for the i-th training example, J can be rewritten as:

$$J_{wse} = \sum_{i=1}^{N} \mathbf{w}_i \left(\mathbf{z}_i - \mathbf{h}_m(\mathbf{v}_i)\right)^2, \tag{5}$$

where wse stands for weighted square error.

The particular equations that govern the minimization of the cost depend on the form of the weak-learners. In this work and in order to minimize the computational cost of the detector, weak learners are implemented by *regression stumps*. A regression stump is defined as follows:

$$\mathbf{h}_m(\mathbf{v}) = a\delta(\mathbf{v}^f > \theta) + b\delta(\mathbf{v}^f \leq \theta) \tag{6}$$

where a y b are the regression parameters, δ is the indicator function, and θ denotes the threshold in the decision. To minimize this cost with respect to the model parameters we proceed as follows: for each candidate feature f, we evaluate all possible thresholds θ and estimate the optimal a and b by minimizing a weighted least squares problem (for more

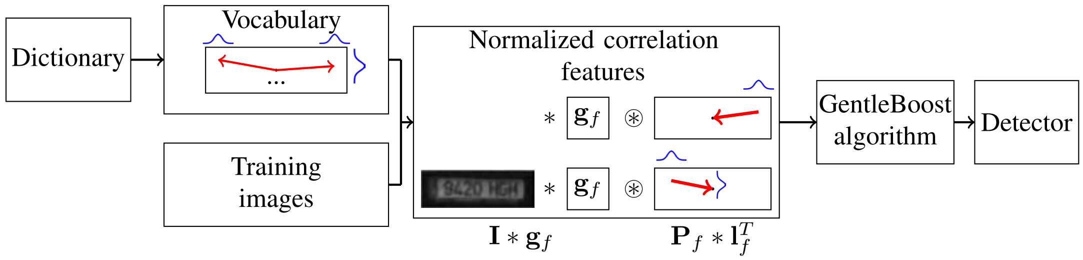


Fig. 2. Block diagram for the baseline detector. The training process is performed with fixed-size license plates.

// Initialization:
$\mathbf{w}_i = 1$ and $\mathbf{H}(\mathbf{v}_i) = 0$
for $m = 1$ **to** M **do**
  // Fit regression stump:
  $\mathbf{h}_m(\mathbf{v}) = a\delta(\mathbf{v}^f > \theta) + b\delta(\mathbf{v}^f \leq \theta)$
  // Update class estimates:
  $\mathbf{H}(\mathbf{v}_i) := \mathbf{H}(\mathbf{v}_i) + \mathbf{h}_m(\mathbf{v}_i)$
  // Update weights for examples:
  $\mathbf{w}_i := \mathbf{w}_i e^{-\mathbf{z}_i \mathbf{h}_m(\mathbf{v}_i)}$
end for

Fig. 3. GentleBoost algorithm for binary classification with regression stumps.

information see [15]). The triplet (a, b, θ) that minimizes the cost is selected and the cost is stored. Finally, the feature that provides the minimum cost is chosen with its corresponding parameters (f, a_f, b_f, θ_f), and the strong classifier is then updated.

The GentleBoost algorithm for our problem is summarized in Figure 3.

B. Application to the LPD problem

The application of this general object detection framework to the particular LPD problem requires two main adaptations: an appropriate selection of training samples, and a non-maxima suppresion at test time.

1) Selection of training samples: the patches that form the visual vocabulary are extracted from size-normalized license plates (from images of the training set) at a rate of 10 patches per license plate. The patch size varies between 9x9 and 15x15 pixels. In order to highlight gradients, inherent to license plate areas, a particular set of filters $\mathbf{g}_f$ are applied to each patch, namely: Sobel filters for vertical gradients, Sobel filters for horizontal gradients and Laplacian filters (orientation-invariant).

The next step is to draw positive and negative samples to train the detector. Since this task has a significant impact on the final system performance, we have developed a specific method for the selection of positive and hard-negative samples. The process is described in the following paragraphs.

First, for each training image we compute a score map $\mathbf{L}$ by aggregating the normalized score produced by all the features in the vocabulary:

$$\mathbf{L} = \frac{1}{D}\sum_{f=1}^{D} \mathbf{v}^f = \frac{1}{D}\sum_{f=1}^{D} (\mathbf{I} * \mathbf{g}_f) \circledast \mathbf{P}_f * \mathbf{l}_f^T. \tag{7}$$

Subsequently, some samples are selected to train the detector. In particular, the local maximum of the score map closest to the center of the license plate is selected as a positive sample, whereas some samples of the background producing high scores are also selected as negatives (i.e., hard negatives). Furthermore, the rate of negative to positive samples has been fixed to 30:1 (i.e., 30 negative samples for each positive one) since the background exhibits much higher degree of variability.

2) Non-maxima suppression: once the detector has been trained, it can be applied to test images, as illustrated in Figure 4. The detector computes the normalized correlation between the filtered test image and the corresponding patch from the vocabulary for every weak classifier. Then, the regression stump associated with the weak classifier is applied and the result added to the boosting margin (see Figure 4b). Subsequently, a threshold is applied to the margin and the pixels exceeding the threshold become candidates to contain a license plate, whereas the rest of the image is discarded. Nevertheless, this method generates multiple detections in close locations all of them coming from the same license plate, thus increasing the false alarm rate. To address this problem, we post-process the thresholded boosting margin by convolving the resulting margin with a Hamming window to group close detections (see Figure 4c). Finally, the center of each resulting cluster is selected as the center for the proposal, as seen in Figure 4d. A filtering process based on the score of the resulting bounding boxes can be applied to discard some of the false alarms.

C. Scale-adaptive part-based model

The first innovation proposed to improve the baseline system is the design of a scale-adaptive part-based model working over a discrete representation of the scale space. This approach contrasts with previous works, in which every license plate was first re-sized to a normalized scale during training, and a pyramid representation of images was then used in test to perform the multi-scale detection. From our point of view, our approach shows some advantages over the baseline system:

- First, the classifier can extract some knowledge about the intra-scale variance present in the database. Hence, the resulting classifier is more robust against the variability in the size of plates due to the distance to the camera. This fact, together with the robustness against illumination and perspective provided by the weak-classifiers, results in a powerful license plate detector.

(a) (b) (c) (d)

Fig. 4. Test process of the detector: (a) original image; (b) image result of the boosting margin; (c) result of the grouping process applied on the thresholded boosting margin; and (d) resulting bounding boxes.

- Second, the detector becomes faster in the test phase. The traditional detectors must search in different scales, which increases linearly the computational time of the test procedure and hinders real-time operation.
- Third, the detector takes advantage of the correlation between adjacent scales, thus improving its performance and the subsequent segmentation process.

The modified features form now the 4-tuple $(\mathbf{g}_f, \mathbf{P}_f, \mathbf{l}_f, s_f)$, where s_f is the discrete scale associated with the f-th feature. Hence, each feature includes the filter, the patch, the location respect to the center of the license plate and its discrete scale. At each iteration, the GentleBoost algorithm selects the feature that provides the highest error reduction (in the whole scale space), as well as its associated scale. Thus, the number of features and their scales are automatically selected by the boosting algorithm from the training data, and the classifier is not restricted to use a manually-fixed number of features for each scale.

Therefore, if a scale is less common in the considered scenario (or little useful for not being correlated with the rest of them), the GentleBoost algorithm will select a small number of features for this scale, thus reducing its impact on the final results. In contrast, the most likely scales (or those which result useful due to their correlation with other scales) will have a greater influence in the detector.

In our approach, the scale space has been divided into four partitions, and the patch size and standard deviations of the location model have been consequently adapted to each scale. The resulting scale space S_j is defined as:

$$S_j = \begin{cases} S_1 & if \quad h < 25 \wedge w < 80 \\ S_2 & if \quad 25 \leq h < 35 \vee 80 \leq w < 100 \\ S_3 & if \quad 35 \leq h < 45 \vee 100 \leq w < 120 \\ S_4 & if \quad 45 \leq h \vee 120 \leq w \end{cases} \tag{8}$$

where h denotes the license plate height and w its width in pixels. The corresponding model parameters for each scale will be introduced and discussed in the following sections.

Finally, in order to improve the segmentation of the license plate, a scale-weighted linear interpolation of the average license plates sizes in each scale has been proposed. Specifically, a score image is computed for every scale by aggregating all the scores coming from the features at each scale and normalizing the results to sum to one. Then, the resulting scores associated with each scale are used as weights for a linear interpolation of the corresponding license plate sizes for each scale, producing in this way a better adapted bounding box. The process is mathematically described in (9):

$$[h'_y, w'_x] = \sum_{j=1}^{4} \frac{\sum_{m \in S_j} \mathbf{h}_m(y,x)}{\sum_m \mathbf{h}_m(y,x)} [\overline{h}_j, \overline{w}_j] \tag{9}$$

where h'_y and w'_x are the dimensions (height and width) of the bounding box located in coordinates y and x (detector proposal) and $\overline{h}_j$ and $\overline{w}_j$ the average height and width for j-th scale.

D. Empirically constrained-deformation part-based model

The second innovation proposed in this paper is the design of a constrained-deformation part-based model for license plates. In our particular scenario, parts of the model are associated with specific details in the plates, and may fall either on the boundaries of the plate (e.g. corners) or in internal areas associated to letters or digits. It is easy to imagine that the former are parts in which the relative location with respect to the plate center remains quite stable for a given scale, whereas the location of patches located in digits or letters is much more variable due to the varying order of characters in licenses.

Consequently, we have designed a method that automatically handles this variable behavior. In particular, the standard deviation of the Gaussian function used for modeling the location of a patch (see $\mathbf{l}_f$ in section III-A1) has been modeled as a combination of two terms: $\boldsymbol{\sigma}_{sc}$, a constant term which only depends on the scale of the feature; and $\boldsymbol{\sigma}_{emp}$, which is learned from the training set and intends to model the variable location of patches with respect to the center of the license plate. In particular, the standard deviation of the patch location is estimated as the average of both terms:

$$\boldsymbol{\sigma}_{loc} = \frac{\boldsymbol{\sigma}_{sc} + \boldsymbol{\sigma}_{emp}}{2} \tag{10}$$

where $\boldsymbol{\sigma}_{emp} = \left(\sigma^y_{emp}, \sigma^x_{emp}\right)$ is the vector containing the empirical deviation along each coordinate and is computed as follows:

- First, for a given feature f, we select the set of locations from every training license plate that show a normalized cross-correlation value higher than 0.8. These locations and their scores are stored in $\mathbf{x}'$, $\mathbf{y}'$ and $\boldsymbol{\alpha}$: the arrays

Fig. 5. Estimation of the standard deviation of the location for a visual word. The original location of the visual word remains constant and the standard deviations (for the two-dimensional Gaussian distribution which models the visual-word location uncertainty) in both x- and y-direction are calculated from the training license plates, by means of a weighted mean of the x and y position of the points, where the normalized correlation is above a threshold α. In this example, the visual word has high variance in x-direction and low variance in y-direction.

containing the x-position, the y-position and the score of the N_s samples with score above the threshold, respectively.

- Second, we estimate each coordinate of the deviation vector from a weighted mean of the relative location of each sample respect to the original location of the feature (x_f, y_f):

$$\sigma^x_{emp}(f) = \frac{1}{\sum \boldsymbol{\alpha}} \sum_k^{N_s} \boldsymbol{\alpha}(k) \cdot (\mathbf{x}'(k) - x_f) \tag{11}$$

$$\sigma^y_{emp}(f) = \frac{1}{\sum \boldsymbol{\alpha}} \sum_k^{N_s} \boldsymbol{\alpha}(k) \cdot (\mathbf{y}'(k) - y_f) \tag{12}$$

- Finally, we obtain $\boldsymbol{\sigma}_{loc}$ as defined in (10).

Figure 5 illustrates the process of calculating the empirical standard deviations for a visual word.

Following this approach, features coming from a visual word with a small variance result in very sharp blobs with high score values affecting to very restricted detection areas. In contrast, features coming from a visual word with a high variance result in flatter blobs with lower scores shared over larger detection areas. Generally, features coming from a visual word with a small variance are preferred by the GentleBoost algorithm because sharp blobs are easier to distinguish from the background, but visual words with higher variance are also useful for the detection process, especially if we apply the grouping process which accumulates the score in nearby areas. If enough features are selected, the detector can properly model the variance of the license plate class as a whole because a representative set of the features are detected in each license plate (and usually not detected in other areas of the image).

Hence, a constrained-deformation part-based model has been proposed. In particular, the standard deviation, $\boldsymbol{\sigma}_{loc}$, is kept within a small range of values (empirically the GentleBoost algorithm works properly with small values while larger values make the precision decrease). Thereby, with the empirically constrained-deformation part-based model the resulting detector is able to adapt to the variations in each feature position whereas the precision remains high.

TABLE I
MAXIMUM AND MINIMUM DIMENSIONS OF THE PLATES IN THE OS DATABASE

Dimension	Minimum	Maximum
Height (pixels)	13	51
Width (pixels)	47	158
Aspect ratio	2.05	5.30

IV. Experiments

The proposed LPD system is assessed in two different datasets in comparison with several algorithms of the state-of-the-art. Before describing the experiments and their results, we will first introduce the datasets and the performance measures, as well as the algorithms selected for the comparative evaluation.

A. Datasets and performance measures

Three datasets have been used in our experiments which will denote OS, Stills [36] and Caltech [37].

The first dataset used in this paper (OS) is composed of 630 images: 246 are used for training and 384 for test. The images have two possible sizes: 1286x986 or 1296x976 pixel (in color JPEG format). Each image can contain several legible license plates. In particular, the number of license plates is 372 for training and 523 for test. The legible license plates have been annotated manually by means of the PASCAL Visual Object Classes software [38], which uses rectangular bounding boxes.

The images have been taken by surveillance cameras in ten different locations under varying lighting conditions (day and night), and usually show the cars from the rear. The size of the license plates changes due to the variable distance from the camera to the cars, as well as their aspect ratio can vary due to perspective. The distance between cameras and legible license plates is between 8 and 25 m. Table I shows the minimum and maximum value for width, height and aspect ratio of the license plates. Figure 6 shows some illustrative examples of images from the database.

The second database (henceforth 'Stills') is available by request [36]. In this case, the images are taken in parkings at a closer distance (less than 5 meters), their size is 640x480 pixels. License plates are labeled and their sizes are very

Fig. 6. Some illustrative examples from the OS database. As can be observed the illumination conditions, the distances from the camera to the cars, and the perspectives are very varied.

Fig. 7. Some examples from Stills [36] (two on the left) and Caltech [37] (two on the right) databases. As can be seen, the scenario is significantly less demanding than that of the OS database.

TABLE II
MAXIMUM AND MINIMUM DIMENSIONS OF THE PLATES IN THE STILLS&CALTECH DATABASE

Dimension	Minimum	Maximum
Height (pixels)	16	36
Width (pixels)	46	112
Aspect ratio	2.26	4.25

similar to each other. The database is divided into a training set (186 images) and a test set (60 images).

The third dataset (henceforth 'Caltech') is publicly available [37]. It is composed of 126 images taken in parkings at a distance of less than 5 meters and their size is 896x572 pixels. The dataset has been manually annotated using PASCAL software [38], and it has been added to 'Stills' test set making a mixed test set of 186 images. In this manner, we can test how the proposed method performs when trained in a database (Stills) and tested on another (Caltech). Table II summarizes the information regarding the size of the license plates for 'Stills' and 'Caltech' databases (Stills&Caltech) and some illustrative examples of images are provided in Figure 7.

For the OS database we have simulated the result of a background subtraction technique to reduce the false alarm rate for all detectors (for the comparison to be fair). In particular, polygonal masks for each one of the 10 distinct scenarios of the database have been built, indicating the image areas that may contain a license plate. If a bounding box does not have at least 60% of its area in this target area, it is discarded.

The performance is calculated in terms of information retrieval measures: the precision-recall (PR) curve and the average precision (AP) for the soft-decision made by the boosting classifiers and the F1 or F-score in (13) for the hard-decision classifiers. The criterion used for considering a detection as a true positive is that its center does not deviate more than 50% of the true center. Additionally, the quality of the segmentation is measured with the Sørensen-Dice coefficient over the pixels of the license plate, as defined in (14), being A the set of pixels corresponding to the ground-truth bounding box and B the set of pixels corresponding to the retrieved bounding box.

$$F1 = \frac{2 \cdot (precision \cdot recall)}{recall + precision} \quad (13)$$

$$SDC = \frac{2 \cdot (A \cap B)}{|A| + |B|} \quad (14)$$

B. Optimization of model parameters

The selection of model parameters is made through a 5-fold cross-validation procedure over the training set. First, the version of the baseline detector [15] adapted to the LPD problem uses filters with sizes between 3x3 and 9x9 pixels (instead of the fixed 3x3 size of the baseline detector). In this way, our proposed approach is suitable for patches with longer gradients, found in larger license plates. The total number of visual words D is 5330 for OS database and 3120 for Stills database: 10 patches x 13 filters x 41 or 24, respectively, dictionary samples. In test, the size of the Hamming window (used in the non-maxima suppression stage to group close detections) is fixed to 50x150 pixels, approximately the size of the largest license plate in the databases.

Table III shows the optimal parameters for the proposed discretization of the scale space. The patch size varies with the scale (the patches should remain representative of the

TABLE III
SUMMARY OF THE PARAMETERS ASSOCIATED WITH EACH OF THE SCALES CONSIDERED

Scale (S_j)	Patch size	$\boldsymbol{\sigma}_{\mathbf{sc}} = [\sigma^{\mathbf{y}}_{\mathbf{sc}}, \sigma^{\mathbf{x}}_{\mathbf{sc}}]$	$[\overline{\mathbf{h}}_{\mathbf{j}}, \overline{\mathbf{w}}_{\mathbf{j}}]$
1	9	[5,5]	[20,70]
2	13	[7,7]	[30,90]
3	17	[9,9]	[40,110]
4	21	[11,11]	[50,130]

license plate content independently of the license plate size). The standard deviation associated with the scale, $\boldsymbol{\sigma}_{sc}$, also varies (larger visual words have larger standard deviation). The average height and width for the scale-weighted linear interpolation method is also shown in Table III.

C. Comparison with the state-of-the-art

The evaluation of the methods has been performed over the test partition of both databases. The parameters of each detector included in the comparison have been previously optimized using the same procedure as for our proposal (see section IV-B).

In our experiments we have considered the following algorithms found in the literature:

- A simple detector based on morphology by Hsieh et al. [13]. It is a versatile detector based on simple closing, opening and smoothing operations.
- A second detector based on morphology by Gou et al. [6]. It is a recent approach which performs a top-hat transformation, Sobel edge detection and noise filtering. Additionally, a Gaussian filtering previous to the top-hat transformation has been added to the process described in [6] because it increases the performance.
- A boosting detector based on Haar features by Dlagnekov [8]. This detector is based on the detector by Viola-Jones [39]. It uses 2400 Haar features (on the original image, its derivatives and variances) and the classifier is trained by means of the Adaboost algorithm on the normalized license plates (with roughly the medium size of the databases, 25x75 pixels). It uses 100 weak-learners for the classification.
- The baseline boosting detector by Torralba et al. [15] (henceforth 'Baseline') with two different configurations: the license plates normalized to the smallest size in the database (15x45 pixels) and to a medium size (30x90 pixels). The multi-scale detection uses 4 scales in the first case (downsampling the original image with a factor of 0.8) or 7 scales in the second case (downsampling the original image with a factor of 0.8 and upsampling it with a factor of 1.2). The detector uses 60 weak-learners.

In addition, two versions of our multi-scale detector have been used in the experiments with the goal of assessing each of the contributions of this paper:

- *Proposed-SA*: the baseline detector including the scale-adaptive part-based model (see III-C).
- *Proposed-SA-EC*: the proposed detector, the baseline version including the scale-adaptive part-based model (see III-C) and the empirically constrained-deformation part-based model (see III-D).

It is worth noting that, in order to provide a fair comparison in terms of computational complexity and memory consumption, the baseline boosting detector and the proposed versions of this paper, have been trained using 60 weak-learners.

The results in terms of detection performance and computational cost in a mid-range computer are shown in Table IV. For the soft-decision boosting detectors, the F-score is calculated from the PR curve by sampling a point with a recall immediately higher than morphology-based detectors recall (in this case, 87.38% for OS database and 84.41% for Stills&Caltech database) and obtaining the corresponding precision. The table also includes the AP for the PR curve for the boosting detectors. The results for the detector by Dlagnekov [8] are not included in the comparison because this approach is not efficient in demanding scenarios.

As can be seen in the table, our approach obtains the best results both in terms of F-score and AP, and notably outperforms the rest of the compared alternatives in the complex OS scenario. Although the morphology detectors are very fast, they rely on vertical edges, which are not discriminant enough in the considered databases (they appear in many other elements of the scenes), thus leading to very poor precisions. Regarding the boosting detectors, we have observed that the Haar features do not capture well relevant patterns over the license plate and, in addition, the resulting patterns are not well associated with fixed and stable locations in the license plate. Our detector, however, finds representative patterns of the license plate at quite stable locations with slight variations with respect to the center of the license plate, which are well modeled by our deformation terms.

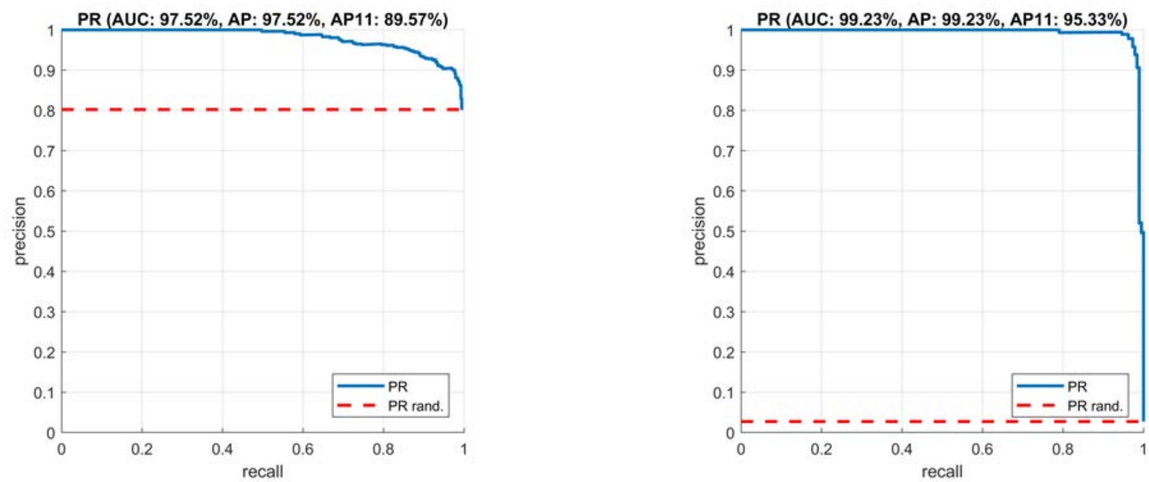
Compared to the baseline detector [15], our proposal does not need to normalize the size of the license plates in training, as it automatically accounts for the variations in scale. We obtain an AP value similar to the baseline method in Stills&Caltech database (an easier task) and a much better performance for the challenging OS scenario: multiple license plates per image, demanding illumination conditions (even night images) and higher variation in license plate sizes. The effect of the contributions is notable in OS database. In particular, the scale-adaptive part-based model achieves a significant increase (10 %) in the AP value, and the empirically constrained-deformation one increases this result in an additional 2%.

In addition, we can see that the results are good in both datasets using the same parametrization, which demonstrates that our approach adapts well to different scenarios. Likewise, the proposed scale-adaptive part-based model reduces the computational cost in comparison with other boosting detectors where multiple scales are searched for based on on the pyramid representation of images.

Figure 9 shows the visual words selected by the GentleBoost algorithm for the detection process. The first visual words have less information and are located in the license plate borders (they are used to detect roughly the license plates in the image) whereas the last ones are more discriminative (they are used

TABLE IV
PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART DETECTORS

Detector	Database	Recall (%)	Precision (%)	F-score (%)	AP (%)	Elapsed time (s)
Hsieh et al. [13]	OS	87.38	52.65	67.71	-	0.08
	Stills&Caltech	84.41	36.68	51.13	-	0.03
Gou et al. [6]	OS	86.23	22.44	35.61	-	0.23
	Stills&Caltech	80.65	21.93	34.48	-	0.11
Dlagnekov [8] (25x75)	OS	-	-	-	-	-
	Stills&Caltech	84.46	31.90	46.31	60.92	23.41
Baseline [15] (15x45)	OS	87.50	51.25	64.64	80.94	16.46
	Stills&Caltech	84.66	100.00	91.69	97.88	5.58
Baseline [15] (30x90)	OS	87.50	53.16	66.14	85.40	77.18
	Stills&Caltech	84.66	74.42	79.21	90.99	23.55
Proposed-SA	OS	87.38	95.61	**91.31**	95.53	8.77
	Stills&Caltech	84.41	100.00	**91.55**	99.14	2.67
Proposed-SA-EC	OS	87.38	94.51	90.81	**97.52**	9.43
	Stills&Caltech	84.41	99.37	91.28	**99.23**	3.16



(a) (b)

Fig. 8. PR curves for the proposed detector for both databases: (a) OS database and (b) Stills&Caltech database.

to locate more precisely the license plates). Some features can be re-used in the detector with different thresholds.

As we have already mentioned, one of the advantages of our approach is that it automatically assigns features to scales during the learning phase, which allows to adapt the visual representation of the license plates to the prior probabilities of the scales (more likely scales will in general be associated to more weak learners in the boosting detector). When analyzing the results, our classifier uses 4, 14, 24 and 18 features for scales 1 to 4, respectively, in the OS database; and 16, 34 and 10 for scales 1 to 3 respectively for Stills database (in this dataset, there are no license plates belonging to scale 4). For the OS database, the assignment of the features is closely related to the “a priori” probability of the license plate size in the database: the number of features are 260, 1300, 2080 and 1690 for scales 1 to 4, respectively. For Stills database the number of features is 1170, 1690 and 260 for scales 1 to 3. In this case the assignation also follows the license plate distribution, but the central scale gains importance because of the correlation with adjacent scales (in this way, some central-scale features can be useful for all the scales).

The segmentation performance is measured through the Sørensen-Dice coefficient (SDC) for each of the compared detectors. In general, although it depends on the type of object, a value of the Sørensen-Dice coefficient above 70% is considered a good segmentation score. For license plates, and due to the fact that the final goal is to read the license, it is also necessary to apply a more restrictive limitation: every digit or character has to be contained in the bounding box. The results are provided in Table V.

The proposed detector outperforms the rest of them in terms of segmentation, except for the baseline one which performs a multi-scale detection. Nonetheless, the loss in segmentation accuracy is negligible in comparison with the increase in detection performance and typically the bounding boxes are large enough to contain all the characters of the license plate. In OS database license plate sizes are associated with their quality (the smaller license plates belong to vehicles distant

(a) (b)

Fig. 9. Visual words selected for the proposed method for both databases: (a) OS database and (b) Stills database.

Fig. 10. Bounding boxes slightly larger than the ground-truth.

TABLE V
SEGMENTATION PERFORMANCE OF THE COMPARED DETECTORS

Detector	Database	SDC (%)
Hsieh et al. [13]	OS	65.82
	Stills&Caltech	61.91
Gou et al. [6]	OS	62.49
	Stills&Caltech	59.07
Dlagnekov [8] (25x75)	OS	-
	Stills&Caltech	78.60
Baseline [15] (15x45)	OS	79.24
	Stills&Caltech	**84.78**
Baseline [15] (30x90)	OS	**79.65**
	Stills&Caltech	46.00
Proposed-SA	OS	73.45
	Stills&Caltech	82.90
Proposed-SA-EC	OS	**74.29**
	Stills&Caltech	**84.13**

to the camera). The visual words with better quality (belonging to larger scales) are more discriminatory to distinguish license plates from the background. Hence, the GentleBoost algorithm tends to include more large-scale visual words in the detector and the interpolation method provides bounding boxes slightly larger than the ground-truth. Some examples can be seen in Figure 10.

Finally, Figure 11 shows a set of images illustrating the detection performance of several detectors in both databases. The selected detectors are the two morphology-based ones, the baseline one, and the one proposed in this paper. The results of our approach outperforms the rest of detectors especially for a demanding task such as the OS one.

D. Error analysis and discussion

In order to provide more insight about the operation of our proposed approach we have analyzed the false alarms produced in both databases. For Stills&Caltech database, false alarms are centered in high-textured areas. Figure 12 shows some of these false alarms. For OS database, false alarms can be divided into three groups: first, high-textured areas as in Stills&Caltech database; second, some illegible license plates that have not been annotated as positives, demonstrating that our detector is coherent and able to detect low-quality license plates. If they had been annotated, the performance of the proposed method would have been even better. The third group of false alarms include some textual areas which can be considered as hard negatives. Figure 13 shows some examples for the three categories.

High-textured false alarms are related with some of the detector's visual words (see Figure 9), which represent high-gradient areas, and textual false alarms only appear in some areas (those whose elements match the inherent structure of the detector). Probably both types of errors might be mitigated adding hard-negatives to the training datasets, i.e., from text databases.

E. Assessment in a large-scale dataset

Finally, we have tested the proposed detector in AOLP dataset [40], a large-scale scenario, to check its performance

Fig. 11. Visual comparison among detectors: MBCS [13], VRBER [6], baseline [15] and proposed (in columns, from left to right). The two first rows refer to the OS database (where baseline [15] normalizes the plate sizes to 15x45 pixels) and the two last rows to Stills&Caltech (where baseline [15] normalizes to 30x90 pixels).

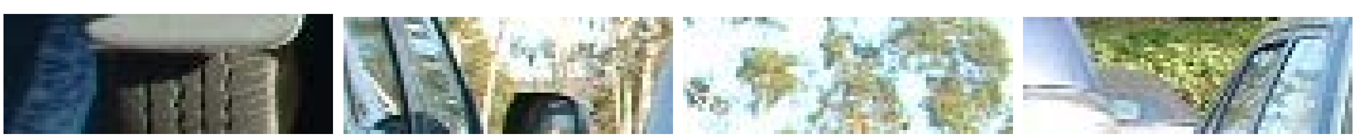


Fig. 12. Some false alarms for Stills&Caltech database.

Fig. 13. Some false alarms for OS database.

TABLE VI
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS FOR AOLP DATASET [40]

Method	F-score (%)
Proposed-SA-EC	**98.98**
Hsu et al. [40]	92.97
Li et al. [27]	97.27
Polishetty et al. [29]	97.80
Xie et al. [25]	**99.47**

in a wider setting. AOLP dataset is composed of 2049 images and is available by request. The images are divided into three subsets with different target applications: Access Control, Traffic Law Enforcement and Road Patrol; with distances between the vehicles and the camera ranging from 5 to 15 meters. In this case, 387 images have been used for training and 1662 for test, as suggested in [40].

The results of the proposed method in comparison with those of state-of-the-art are shown in Table VI. Our proposed approach outperforms the one reported on [40], even though ours shares parameters among the three scenarios and the previous OS and Stills&Caltech datasets described in section IV-B, while theirs adapts its parameters to each scenario. Furthermore, the results of our detector compare well to those recently reported with CNNs [27], [29], [25], which have been specifically trained and designed for this type of scenarios, and therefore constitute the state-of-the-art on it. Although it is noteworthy that AOLP does not pose any challenge with respect to the multiple scales in license plates, the focus of our paper, its large size allows us to assess the good generalization capability of our method.

V. CONCLUSION

A robust approach for vehicle license plate detection based on boosting and part-based models has been proposed in this paper. A set of filtered visual words have been extracted from a subset of images of the training set and the boosting classifier has been trained relying on the normalized correlation of these visual words with training images.

The proposed system has been built on a general object detector [15], which has been adapted to the LPD problem. We have proposed two fundamental extensions over the baseline which constitute the main contributions of the paper. First, the design of a scale-adaptive part-based model which discretizes the scale space, adapts the selection of features to the most prominent scales, avoids the need of multi-scale testing and, furthermore, improves the segmentation process. Second, a constrained-deformation part-based model allows to adapt to varying deformation in local features and makes the system capable to cope with quite diverse scenarios.

The proposed system has been assessed on three different databases, which we have called OS, Stills&Caltech [36], [37] and AOLP [40]. OS dataset contains Spanish license plates, Stills&Caltech dataset contains plates from USA and AOLP one contains plates from Taiwan. The detection performance in terms of AP is over 97% in all cases, which reveals that the proposed method is effective in a wide variety of scenarios.

The proposed system has been compared with several methods of the state-of-the-art, with very favourable results in terms of detection performance. In particular, our proposal is clearly superior to all the compared methods, from simple and fast morphology-based methods [6], [13] to complex and more computationally demanding boosting-based methods [8]. While the proposed method shows similar results to those of the best of the compared methods for Stills&Caltech database (the easiest task), it yields much better performance than any other compared method when a more challenging scenario is addressed (OS dataset). In particular, the proposed method reaches an AP of 97.52% vs. 85.40% achieved by the best of the compared methods. Moreover, the proposed system has been trained on AOLP dataset (a large-scale dataset) obtaining competitive performance when compared with recent CNN-based methods.

The proposed system does not rely on gradients, which are not high enough in many practical scenarios due to the presence of noise, the lack of proper illumination, and blurring owing to distant license plates. Furthermore, the number of weak-learners has been reduced and the necessity to look for the plates at several scales has been avoid to make the system more efficient with the discretization of the scale space. Finally, the proposed part-based model increases both the recall and the precision in demanding scenarios.

However, the developed method can still be improved. Its application in a real environment requires additional training to comprise all the possible situations, especially those whose difficulty lies in illumination conditions (night images, artifacts due to non uniform illumination, etc.). Furthermore, some additional information about morphological structures can be considered to enhance the segmentation process.

ACKNOWLEDGMENT

The authors would like to thank to Omnivisión Seguridad, Louka Dlagnekov, Serge Belongie and Gee-Sern Hsu for supplying the images used for this research.

REFERENCES

[1] H. Caner, H. S. Gecim, and A. Z. Alkar, "Efficient embedded neural-network-based license plate recognition system," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 2675–2683, Sept. 2008.

[2] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 377–391, Sept. 2008.

[3] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi, "An algorithm for license plate recognition applied to intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 830–845, Sept. 2011.

[4] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, Feb. 2013.

[5] C. N. E. Anagnostopoulos, "License plate recognition: A brief tutorial," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 1, pp. 59–67, 2014.

[6] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted boltzmann machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1096–1107, Apr. 2016.

[7] S.-L. Chang, L.-S. Chen, Y.-C. Chung, and S.-W. Chen, "Automatic license plate recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 1, pp. 42–53, Mar. 2004.

[8] L. Dlagnekov, "License plate detection using adaboost," Department of Computer Science and Engineering, UC San Diego, Tech. Rep., 2004.

[9] A. H. Ashtari, M. J. Nordin, and M. Fathy, "An iranian license plate recognition system based on color features," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1690–1705, Aug. 2014.
[10] T. Ajanthan, P. Kamalaruban, and R. Rodrigo, "Automatic number plate recognition in low quality videos," in *Proc. 8th Int. Conf. on Ind. and Inform. Syst.*, Dec. 2013, pp. 566–571.
[11] C.-C. Chen and J.-W. Hsieh, "License plate recognition from low-quality videos," in *Proc. IAPR Conf. on Mach. Vision Applicat.*, May 2007, pp. 122–125.
[12] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu, and N. Komodakis, "A robust and efficient approach to license plate detection," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1102–1114, Mar. 2017.
[13] J.-W. Hsieh, S.-H. Yu, and Y.-S. Chen, "Morphology-based license plate detection from complex scenes," in *Proc. 16th Int. Conf. on Pattern Recognition*, vol. 3, 2002, pp. 176–179.
[14] A. C. Roy, M. K. Hossen, and D. Nag, "License plate detection and character recognition system for commercial vehicles based on morphological approach and template matching," in *Proc. 3rd Int. Conf. on Elect. Eng. and Inform. Commun. Technology*, Sept. 2016, pp. 1–6.
[15] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 854–869, May 2007.
[16] L. Zheng, X. He, B. Samali, and L. T. Yang, "An algorithm for accuracy enhancement of license plate recognition," *J. of Comput. and Syst. Sci.*, vol. 79, no. 2, pp. 245 – 255, 2013.
[17] T. T. Nguyen and T. T. Nguyen, "A real time license plate detection system based on boosting learning algorithm," in *Proc. 5th Int. Congr. on Image and Signal Process.*, Oct 2012, pp. 819–823.
[18] W. Wang, Q. Jiang, X. Zhou, and W. Wan, "Car license plate detection based on MSER," in *Proc. 2011 Int. Conf. on Consumer Electron., Commun. and Networks (CECNet)*, Apr. 2011, pp. 3973–3976.
[19] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
[20] B. A. Fomani and A. Shahbahrami, "License plate detection using adaptive morphological closing and local adaptive thresholding," in *Proc. 3rd Int. Conf. on Pattern Recognition and Image Anal. (IPRIA)*, Apr. 2017, pp. 146–150.
[21] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. 6th Int. Conf. on Comput. Vision*, Jan. 1998, pp. 555–562.
[22] A. Elbamby, E. E. Hemayed, D. Helal, and M. Rehan, "Real-time automatic multi-style license plate detection in videos," in *Proc. 12th International Computer Engineering Conference (ICENCO)*, Dec. 2016, pp. 148–153.
[23] C. T. Hsieh, L.-C. Chang, K. M. Hung, and H.-C. Huang, "A real-time mobile vehicle license plate detection and recognition for vehicle monitoring and management," in *Proc. 2009 Joint Conf. on Pervasive Computing (JCPC)*, Dec 2009, pp. 197–202.
[24] H. He, Z. Shao, and J. Tan, "Recognition of car makes and models from a single traffic-camera image," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3182–3192, Dec. 2015.
[25] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, Feb. 2018.
[26] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, "Segmentation- and annotation-free license plate recognition with deep localization and failure identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2351–2363, Sept. 2017.
[27] H. Li and C. Shen, "Reading car license plates using deep convolutional neural networks and LSTMs," *CoRR*, vol. abs/1601.05610, 2016. [Online]. Available: http://arxiv.org/abs/1601.05610
[28] C. Gerber and M. Chung, "Number plate detection with a multi-convolutional neural network approach with optical character recognition for mobile devices," *J. of Inform. Process.*, vol. 12, no. 1, pp. 100–108, Mar. 2016.
[29] R. Polishetty, M. Roopaei, and P. Rad, "A next-generation secure cloud-based deep learning license plate recognition for smart cities," in *Proc. 15th IEEE Int. Conf. on Mach. Learning and Applications (ICMLA)*, Dec. 2016, pp. 286–293.
[30] S. G. Kim, H. G. Jeon, and H. I. Koo, "Deep-learning-based license plate detection method using vehicle region extraction," *Electron. Lett.*, vol. 53, no. 15, pp. 1034–1036, 2017.
[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, pp. 91–99.
[32] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. of Comput. Vision*, vol. 104, no. 2, pp. 154–171, Sep. 2013.
[33] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Ann. of Statistics*, vol. 28, no. 2, pp. 337–407, Apr. 2000.
[34] R. E. Schapire, "A brief introduction to boosting," in *Proc. 16th Int. Joint Conf. on Artificial Intell.* Morgan Kaufmann Publishers Inc., 1999, pp. 1401–1406.
[35] J. Sun, D. Cui, D. Gu, H. Cai, and G. Liu, "Empirical analysis of adaboost algorithms on license plate detection," in *Proc. 2009 Int. Conf. on Mechatronics and Automation*, Aug. 2009, pp. 3497–3502.
[36] L. Dlagnekov and S. Belongie, "UCSD/Calit2 car license plate, make and model database," http://vision.ucsd.edu/belongie-grp/research/carRec/car_data.html, 2006.
[37] M. Weber, "Caltech cars 1999 dataset," http://www.vision.caltech.edu/Image_Datasets/cars_markus/cars_markus.tar, 1999.
[38] "Annotation software PASCAL visual object classes," http://host.robots.ox.ac.uk/pascal/VOC/PAScode.tar.gz, 2012.
[39] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Comput. Soc. Conf. on Comput. Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, pp. 511–518.
[40] G. S. Hsu, J. C. Chen, and Y. Z. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 552–561, Feb 2013.

Miguel Molina-Moreno was born in Rus, Jaén, Spain in 1993. He received the Telecommunications Engineering Degree from Universidad de Granada, Granada, Spain in 2015.

Since 2015 he has been studying the M. Sc. in Telecommunications Engineering and Multimedia and Communications as well as working as a Research Assistant in the Multimedia Processing Group in Universidad Carlos III de Madrid, Madrid, Spain. His research interests include computer vision, object detection, signal processing and machine learning.

Iván González-Díaz received the Telecommunications Engineering degree from Universidad de Valladolid, Valladolid, Spain, in 1999, the M.Sc. and Ph.D. degree from Universidad Carlos III de Madrid, Madrid, Spain, in 2007 and 2011, respectively. After holding a postdoc position in the Laboratoire Bordelais de Recherche en Informatique at the University Bordeaux, he currently works as a Visiting Professor at the Signal Theory and Communications Department in Universidad Carlos III de Madrid.

His primary research interests include object recognition, category-based image segmentation, scene understanding and content-based image and video retrieval systems. In these fields, he is co-author of several papers in prestigious international journals, two chapters in international books and a few papers in revised international conferences.

Fernando Díaz-de-María (M'97) received the Telecommunication Engineering degree and the Ph.D. degree from the Universidad Politécnica de Madrid, Madrid, Spain, in 1991 and 1996, respectively. From Oct. 1991 to Oct. 1995, he worked as a Lecturer at Universidad de Cantabria, Santander, Spain. From Nov. 1995 to Sep. 1996, he worked as an Assistant Professor at Universidad Politécnica de Madrid. Since October 1996 to December 2017, he was an Associate Professor in the Department of Signal Processing and Communications, Universidad Carlos III de Madrid (UC3M), Madrid, Spain, and since January 2018 he has been a Professor in the same Department.

More than 10 Ph.D. students have completed their Ph.D. thesis under his supervision at UC3M and his work has led to more than 100 publications in international conferences and journals. His current research interests include image and video processing and computer vision. He has led numerous projects and contracts in these fields. He was the recipient of the Best Ph.D. Dissertation in GSM Applications Award (Official Telecommunication Engineer Association) in 1996, and the ICIP 2011 Best Paper Award.