Road-Sign Detection and Recognition Based on Support Vector Machines

Saturnino Maldonado-Bascón, *Member, IEEE*, Sergio Lafuente-Arroyo, Pedro Gil-Jiménez, Hilario Gómez-Moreno, *Member, IEEE*, and Francisco López-Ferreras

***Abstract*—This paper presents an automatic road-sign detection and recognition system based on support vector machines (SVMs). In automatic traffic-sign maintenance and in a visual driver-assistance system, road-sign detection and recognition are two of the most important functions. Our system is able to detect and recognize circular, rectangular, triangular, and octagonal signs and, hence, covers all existing Spanish traffic-sign shapes. Road signs provide drivers important information and help them to drive more safely and more easily by guiding and warning them and thus regulating their actions. The proposed recognition system is based on the generalization properties of SVMs. The system consists of three stages: 1) segmentation according to the color of the pixel; 2) traffic-sign detection by shape classification using linear SVMs; and 3) content recognition based on Gaussian-kernel SVMs. Because of the used segmentation stage by red, blue, yellow, white, or combinations of these colors, all traffic signs can be detected, and some of them can be detected by several colors. Results show a high success rate and a very low amount of false positives in the final recognition stage. From these results, we can conclude that the proposed algorithm is invariant to translation, rotation, scale, and, in many situations, even to partial occlusions.**

***Index Terms*—Classification, detection, hue, hue saturation intensity (HSI), road sign, support vector machines (SVMs).**

I. Introduction

TRAFFIC-SIGN detection and recognition have been an important issue for research recently. Traffic signs have a dual role: First, they regulate the traffic and, second, indicate the state of the road, guiding and warning drivers and pedestrians. These signs can be classified according to their color and shape, and both these characteristics constitute their content, as shown in Table I. The visibility of traffic signs is crucial for the drivers' safety. For example, very serious accidents happen when drivers do not notice a stop sign, as pointed out in [1]. Of course, many other accidents are not related to traffic signs and are due to factors such as the psychological state of drivers. The causes for accidents that are related to traffic signs may be occlusion or partial occlusion of the sign, deterioration of the sign, or possible distraction of the driver. This work is twofold: First, the algorithm that is described here can be used to evaluate the signaling of the road for maintenance purposes, and second, in a future application, it can be used for driver-assistance systems.

TABLE I
MEANING OF SPANISH TRAFFIC SIGNS ACCORDING TO THE COLOR AND SHAPE

Color	Shape	Meaning
Red Rim	Circle	Prohibition
Red Rim (Up)	Triangle	Danger
Red Rim (Down)	Triangle	Yield
Red	Octagonal	Stop
Blue	Square	Recommendation
Blue	Circle	Obligation
White	Circle	End of prohibition
Yellow	Circle	End of prohibition (construction)

It is interesting to pay attention to some of the common problems that are involved in traffic-sign detection. These problems are analyzed, and a test set has been reported in [2] and [3]. The first problem to be overcome is caused by the variable lighting conditions of the scene in a natural environment. These different conditions are brought about by changes in the weather (producing bright spots and shadows), the time of the day or night, and the state of the road sign itself subject to deterioration or vandalism. Another problem to surpass is the possible rotation of the signs. Although the perfect position for a road sign is perpendicular to the trajectory of the vehicle, many times, the sign is not positioned that way. Therefore, an automatic system must be able to detect signs in many positions and, hence, must be invariant to rotation and translation. The next problem is related to traffic-sign size because we find signs of different dimensions, although officially, there are only three normal sizes for nonurban environments. Due to the method that we follow to capture the signs, we get an image sequence as we approach the sign. Thus, our aim is to implement an algorithm to detect the sign as soon as it becomes visible. For this reason, we cannot accept a system that only detects signs of a specific size. The last problem that we describe here has been mentioned previously. Occlusions often occur because objects such as trees, other signs, or vehicles can reduce the visibility of the signs and, in conclusion, make the detection system fail. In this paper, we overcome these problems with a high success rate.

Finally, the number of different signs is quite large. They are summarized in Fig. 1(a)–(f), where all signs that are employed in Spain are illustrated. Due to this amount, we have to

Manuscript received July 7, 2005; revised March 2, 2006, August 28, 2006, and December 1, 2006. This work was supported by the Ministerio de Educación y Ciencia de España under Project TEC2004/03511/TCM. The Associate Editor for this paper was A. Zelinsky.

The authors are with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain (e-mail: saturnino.maldonado@uah.es; sergio.lafuente@uah.es; pedro.gil@uah.es; hilario.gomez@uah.es; francisco.lopez@uah.es).

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.2007.895311

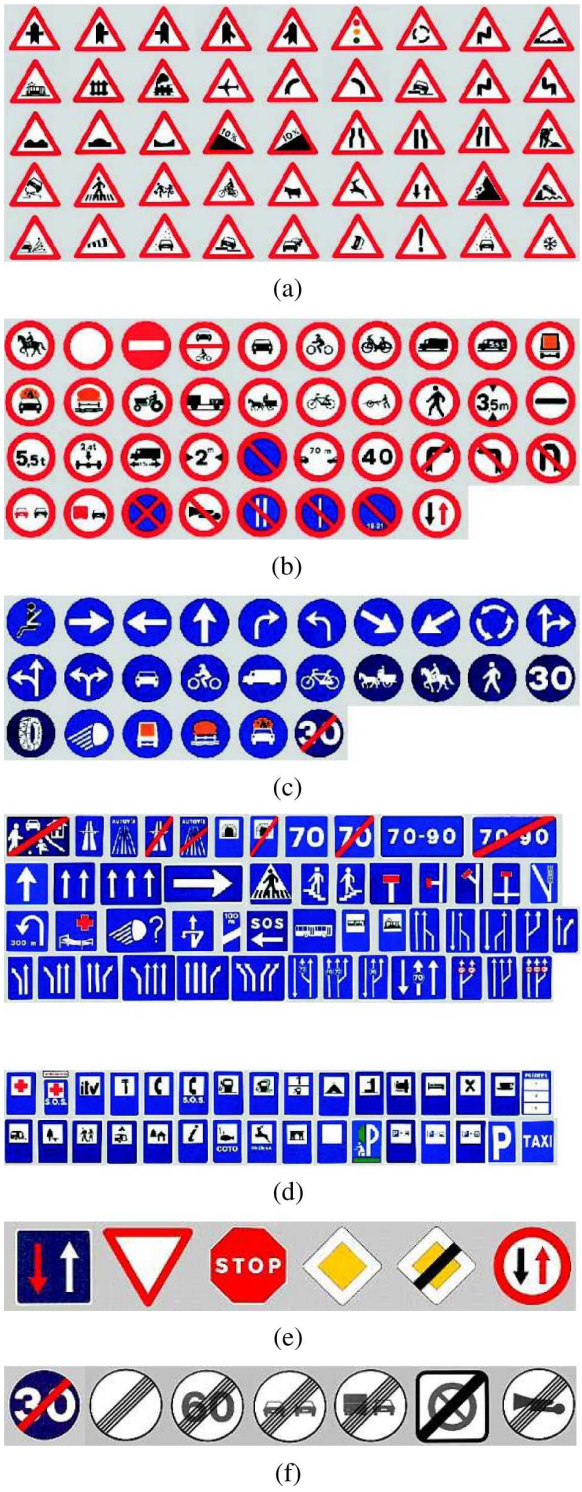


Fig. 1. Traffic-sign database. (a) Danger. (b) Prohibition. (c) Obligation. (d) Warning and information. (e) Priority. (f) End of prohibition.

develop strategies to divide the recognition process into different groups.

The organization of this paper is given as follows: Section II focuses on previous research works. We will present first the traffic-sign detection process and then the traffic-sign recognition of the inner area in correspondence to the previously detected sign, with respect to the most relevant techniques that are developed for this purpose.

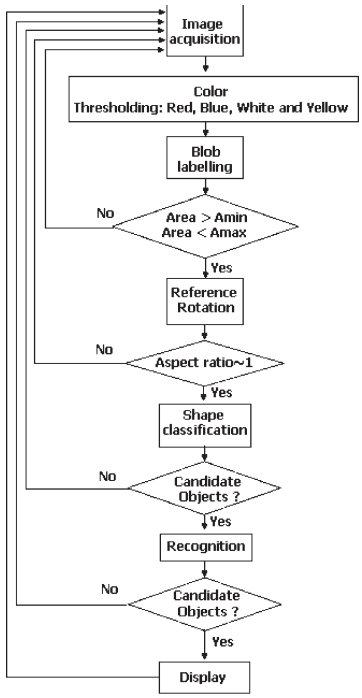


Fig. 2. Algorithm description.

The structure of Section III consists of an exhaustive description of detection and recognition stages, which follows the algorithm shown in Fig. 2. The complete process is triggered by color segmentation of the frame, where the system will search objects with similar colors as traffic signs, i.e., red, blue, white, and yellow, as presented in Section III-A. Once all objects representing possible candidates have been extracted, some features such as size or aspect ratio are analyzed using a table of all the geometric shapes in which traffic signs are designed. Therefore, only the objects that are not discarded in this selection step are analyzed in posterior modules. As the number of different traffic signs is quite large, shape classification is performed prior to the recognition module using the distance to borders (DtBs) as input vectors (as presented in Section III-B) in order to improve computation time in the next stage of recognition. As traffic signs can appear at many different angles to the camera's optical axis, each blob of interest is rotated until they are all aligned in the same way before the classification process begins. Every image is processed in the same way; if no suitable objects that correlate in either color and geometric properties are found, another frame is analyzed. Section III-C deals with the recognition module. Both modules, i.e., shape classification and traffic-sign recognition, are implemented with SVMs, which is a novel technique in this field. In the complete system, the possible outputs of both modules are determined by color. By establishing the color first, only some geometric shapes and only a subset of the traffic-sign database are then liable to be considered in the classification and recognition stages. Therefore, those objects with similar colors as traffic signs (for example, cars and buildings), which hence represent noisy objects for our system, are rejected with high probability in one of these three selective stages: 1) geometric feature selection; 2) shape classification; and 3) recognition of the inner area.

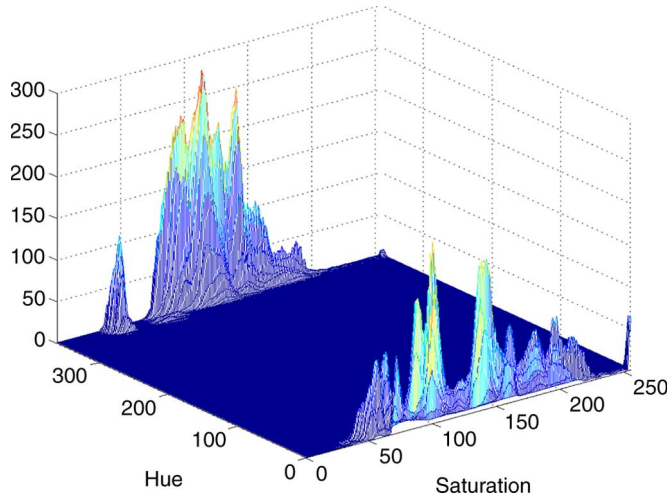


Fig. 3. Hue–saturation histogram for red signs.

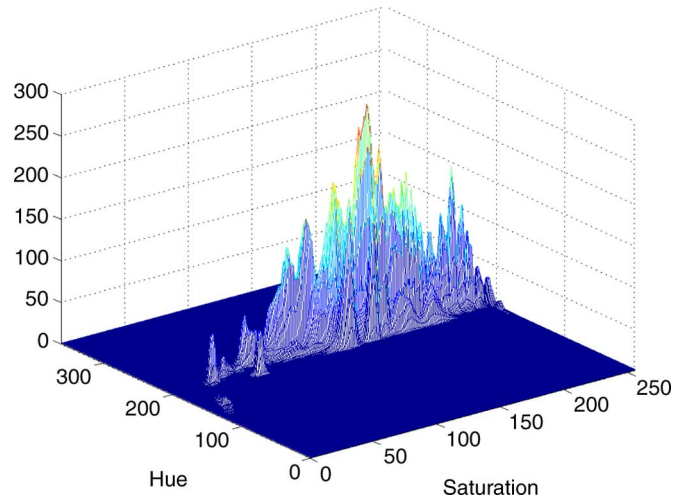


Fig. 4. Hue–saturation histogram for blue signs.

Results that are illustrated in Section IV will confirm the robustness of the proposed algorithm. Finally, conclusions will be presented in Section V.

II. State of the Art

Traffic-sign recognition algorithms are divided, in most cases, into two stages: 1) detection and 2) recognition.

In many works, the first block of the detection system consists of a segmentation stage by thresholding with a given color space to extract the sign color from the image. Direct thresholding over the red green blue (RGB) space is seldom used because it is very sensitive to lighting changes. Thus, in [4], a color ratio between the intensity of the specific RGB color components and the sum of intensity of RGB is used to detect strong colors in the image. A different relation between the RGB components is employed in [5], where one component is taken as a reference. Although other spaces are used, for example, the YUV system is considered in [6] to detect blue rectangular signs, the most frequently employed space is the hue saturation intensity (HSI) system because color information, which is encoded with hue and saturation components, presents low variations for objects of interest with a similar color. In [7], proper thresholds on hue and saturation bands are fixed to extract the red and blue colors. In [8], a nonlinear transformation over hue and saturation is employed to enhance the desired colors in the image (red and blue) using two lookup tables for every color for which we are looking. A similarity measurement between the hue component and the previously stored hue values of particular colors in road signs is calculated in [9], and this measurement is fed into a perceptual analyzer that is implemented by a neuronal network.

Nevertheless, there are some works where thresholding is not applied directly using a specific color space. Thresholding by looking for chromatic and achromatic colors is applied in [10] by a simple vector filter (SVF). The SVF has characteristics that can extract the specific color and eliminate all the outlines at the same time. Normalized error between the luminance component and RGB components is obtained in [8], computing an energy function in order to identify the inner area of the sign.

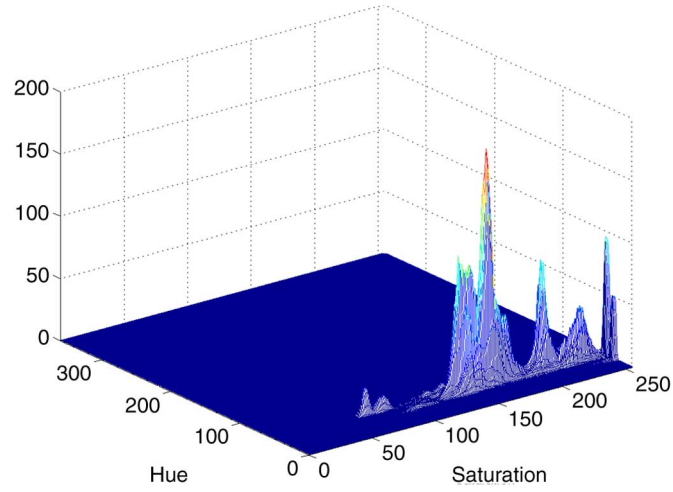


Fig. 5. Hue–saturation histogram for yellow signs.

Other methods for detection have been developed based on edge detection. Many robust-shaped detectors such as Hough circle transform are slow to compute over large images. Nevertheless, some recent works (see [11] and [12]) have implemented a fast algorithm based on radial symmetry that is adapted to triangular, square, diamond, octagonal, and circular shapes. It operates on the gradient of a grayscale image and exploits the nature of shapes that vote a center point for circular signs and a line of votes in the case of regular polygons. The main advantage of this method is that it is able to run in real time. As it detects shapes based on edges, the algorithm is robust to changing illumination.

In [13], a technique based on genetic algorithms is proposed to recognize circular traffic signs by using only the brightness of an input image, which is obtained in the form of a binary image with the help of a smoothing and a Laplacian filter. In [14], two neural networks (NNs) are developed to extract color and shape features. Both these features are applied to a hue component and a gradient image, where the centers of specific color regions and the centers of certain fixed shapes are given.

In [15], the Bayes classifier chooses a color class based on the maximum conditional probability of each color, where equal *a priori* probability for each of the five colors (red, yellow, green, blue, and white) is assumed.

Once the candidate regions have been separated from the image, some research has gone into classifying the candidate signs according to their shape. The classification criterion can be based on the idea of dividing the possible groups of traffic signs by color and shape, such as red circular, red triangular, and blue circular. In [5], different methods using the extracted corners are applied to the traffic-sign shapes, and in [14], color extraction is complemented with shape features using two NNs. Another method is used in [15]. After segmentation using a Bayes classifier, the scale-invariant feature transform that is presented in [16] is applied after the key points, which are going to represent each object that have been determined.

In the recognition stage, an NN is used for the classification, following the Adaptive Resonance Theory paradigm in [17]. In [6], the identification of signs is carried out by a normalized correlation-based pattern matching using a traffic-sign database. In [18], the proposed sign recognition system consists of a nonlinear correlator. The scene and the reference pattern are both Fourier transformed and nonlinearly modified. The correlation plane between the input and the reference signs is obtained by the inverse Fourier transform. In [19], the recognition is done using matching pursuit (MP) in two processes: 1) training and 2) test. The training process finds a set of the best MP filter bases for each road sign. The testing process projects the input unknown road sign to a different set of MP filter bases to find the best match.

Instead of recognition of road-sign symbols, a framework for detecting text on road signs from video is presented in [20]. The system finds a number of discriminative feature points, which are easy to track, in the current video using the detector of Shi and Tomasi.

Although many works are applied in a single-frame way, in [14], the detection system is sensitive to some particular image size, and a Kalman filter is used to track a sign through the frames until it is sufficiently large to be recognized as a specific standard sign. A more recent work [9] presents an automatic road-sign detection and recognition system that is based on a computational model of human visual recognition processing. The system consists of three major components: 1) sensory; 2) perceptual; and 3) conceptual analyzers. The sensory analyzer extracts the spatial and temporal information of interest from video sequences. The extracted information then serves as the input stimuli to a spatiotemporal attentional NN in the perceptual analyzer. If the stimulation continues, the focuses of attention will be established in the NNs where the used features indicate the color horizontal projection of the road sign; they indicate good results because of the tracking system although they consider that road signs are always composed of a color rim with a black/white interior. The detected sign is normalized and correlated with all of the prototypes in [8]; a horizontal and vertical displacement of ± 3 pixels is allowed.

In order to reduce the computation time, in [19], color is used as *a priori* information, and the possible road location is limited to certain designated regions.

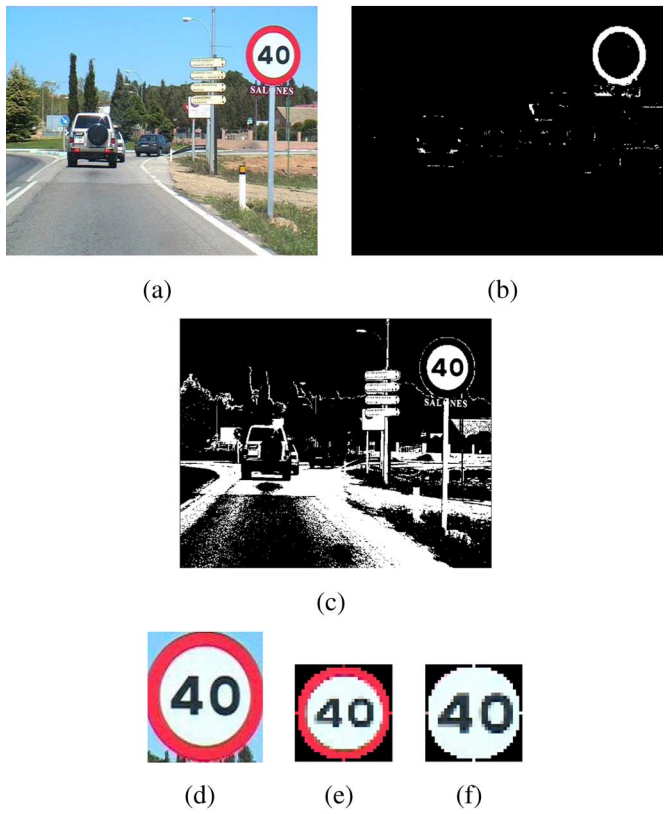


Fig. 6. Sign model as a contribution of two parts. (a) Original image. (b) and (c) Segmentation masks by red and achromatic colors. (d) Extracted sign. (e) Outer and (f) inner regions of normalized sizes.

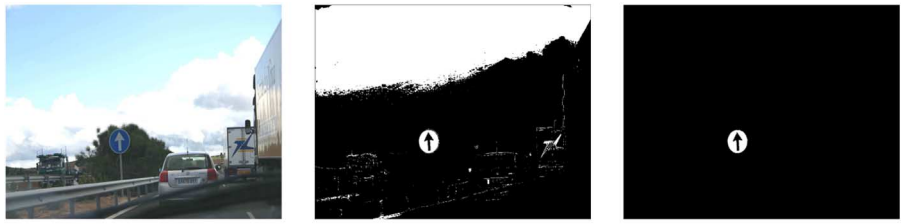
III. System Overview

In this paper, we present a system for detection and recognition of traffic signs that has been successfully applied to Spanish traffic signs. The detection and recognition system consists of three stages.

1) Segmentation: Candidate blobs are extracted from the input image by thresholding using HSI color space for chromatic signs. At the same time, white signs are detected with the help of an achromatic decomposition.
2) Shape classification: Blobs that are obtained from segmentation are classified in this stage using linear SVMs. According to the color that has been used in the segmentation, only some given shapes are possible. For example, signs that are segmented using the red clues can be circular, triangular, or octagonal.
3) Recognition: The recognition process is based on SVMs with Gaussian kernels. Different SVMs are used for each color and shape classification.

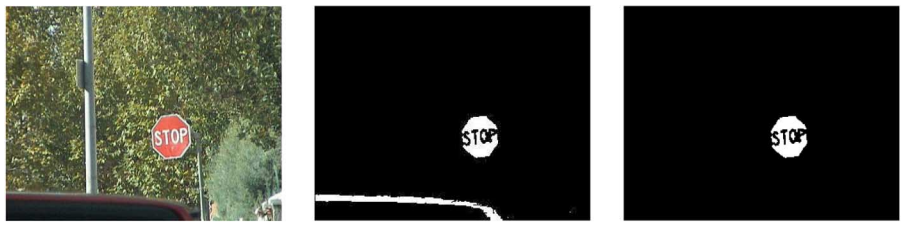
A. Segmentation

As pointed out in Section II, different color spaces have been used to segment traffic signs in outdoor images. The difficulties that we encounter in this image segmentation are related to illumination changes and possible deterioration of the signs. We believe that the hue and saturation components of the HSI space are sufficient to isolate traffic signs in a scene working with fixed thresholds. To obtain these thresholds, we have built



(a) (b) (c)

Fig. 7. Segmentation results.



(a) (b) (c)

Fig. 8. Segmentation results.

the histograms of hue and saturation for red, blue, and yellow of manually segmented signs, as shown in Figs. 3–5. Here, the hue and saturation components take values ranging from 0 to 360 and from 0 to 255, respectively. At this point, we must consider that the response to varying wavelength and intensity of standard imaging devices is nonlinear and interdependent [21]. Due to this reason, the images from which we analyze the hue and saturation components of traffic signs have been taken under different weather and lighting conditions using several cameras with different configurations in order to get well-suited thresholds.

Unfortunately, the hue and saturation components do not contain enough information to segment white signs. The image's achromatic decomposition then helps to detect white signs in a way similar to the method that is described in [10] according to

$$f(R,G,B) = \frac{(|R-G| + |G-B| + |B-R|)}{3D} \qquad (1)$$

where R, G, and B represent the brightness of the respective color, and D is the degree of extraction of an achromatic color, and in our case, we get the best segmentation results by setting D to 20. An $f(R,G,B)$ of less than 1 represents achromatic colors, and an $f(R,G,B)$ of greater than 1 represents chromatic colors.

After the segmentation stage, image pixels may belong to any of the four color categories, i.e., red, blue, yellow, and/or white, and are grouped together as connected components called blobs. Most common signs present a red rim and an inner white region [see Fig. 1(a), (b), and (e)], except when the road is under construction, where the inner area is yellow. This characteristic led us to consider the signs as a possible sum of two contributions corresponding to their chromatic and achromatic segmentation masks, where both parts (see Fig. 6) are processed independently in the complete system.

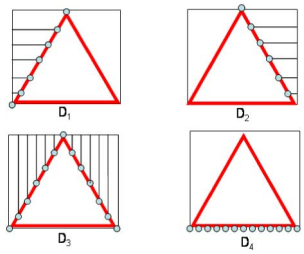


Fig. 9. DtBs for a triangular shape.

The segmentation of candidate traffic signs in scenes that are taken at night is similar to that of the ones that are captured during daylight, except for one consideration: White signs are not well isolated at night by the aforementioned achromatic decomposition. This is due to the color that we perceive from a reflective traffic sign when it is illuminated by the vehicle's headlamp. For this reason, achromatic signs were segmented at night by the hue and saturation components, taking into account that the distribution of the hue components is, in this case, so similar to that of the same component for yellow signs. The difference between both colors, i.e., white and yellow, at night is given by saturation.

All candidate blobs are analyzed in a selection process, and some of them are discarded according to their size or aspect ratio because, given a camera of known focal length, we can impose some limits on the size of objects that we are interested in, i.e., small blobs and big blobs are rejected as noise and noninterest objects, respectively. The limits for both criteria, i.e., size and aspect ratio, were empirically derived based on road images. Thus, the thresholds for the size criterion are fixed at specific percentages with respect to the minor dimension of the images to be analyzed, and only those objects whose dimensions are delimited between one 20th and two thirds will be processed. On the other hand, objects with an aspect ratio of greater than 1.9 and less than 1/1.9 are rejected. Once the segmentation process is completed, we obtain the blobs of

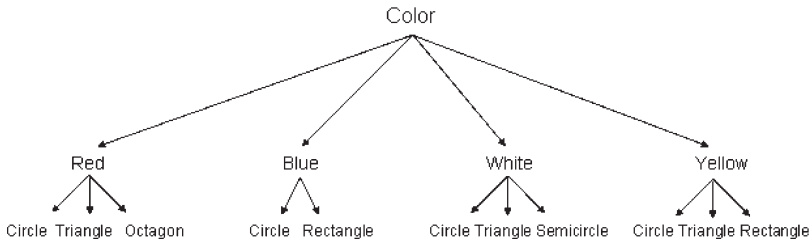


Fig. 10. Tree structure of the classification. The structure shows how the segmentation color determines the possible shapes of the BoI.

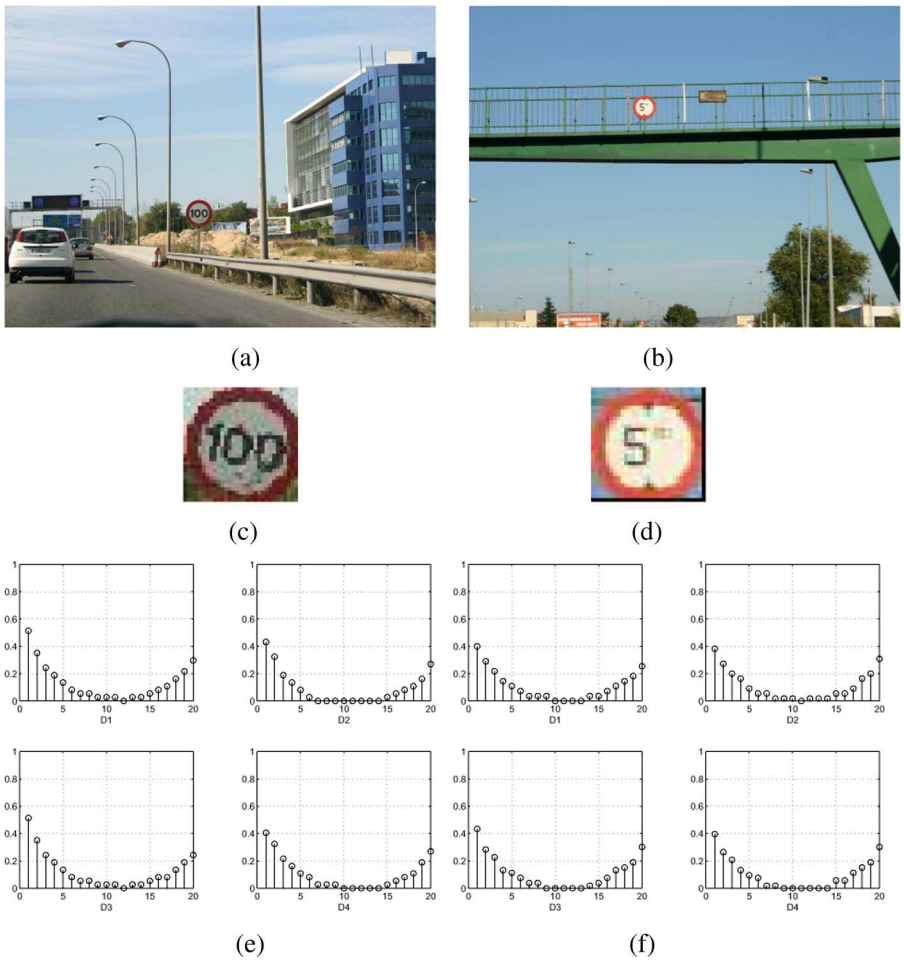


Fig. 11. Translation invariance. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

interest (BoI) or, in other words, possible traffic signs. Since traffic signs are regular polygons or circles, their corresponding bounding box is a rectangle whose aspect ratio is greater than 1 when the camera's optical axis is not perpendicular to the sign, producing a perspective distortion.

Figs. 7 and 8 show the original, segmented, and BoI images for two images of our test set. In Fig. 7, some small blobs and a big blob corresponding to the sky have been removed. In Fig. 8, a blob corresponding to the upper part of a car has been removed for its unsuitable aspect ratio. As we mentioned before, all traffic signs are regular polygons, and their ideal position is perpendicular to the direction of driving. Nevertheless, signs do not always appear in the ideal position, and because of that, each candidate blob in the image is rotated in our system to a reference position before the shape classification in order to obtain a rotation-invariant method.

B. BoI Shape Classification

The blobs that were obtained from the segmentation stage are classified in this stage according to their shape using linear SVMs. SVMs were introduced by Vapnik [22], [23], and some extensive introductions about SVMs can be found in [24] and [25]. The formulation of SVMs deals with structural risk minimization (SRM). SRM minimizes an upper bound on the Vapnik Chervonenkis dimension, and it clearly differs from empirical risk minimization, which minimizes the error on the training data. Although SVMs were developed for and have

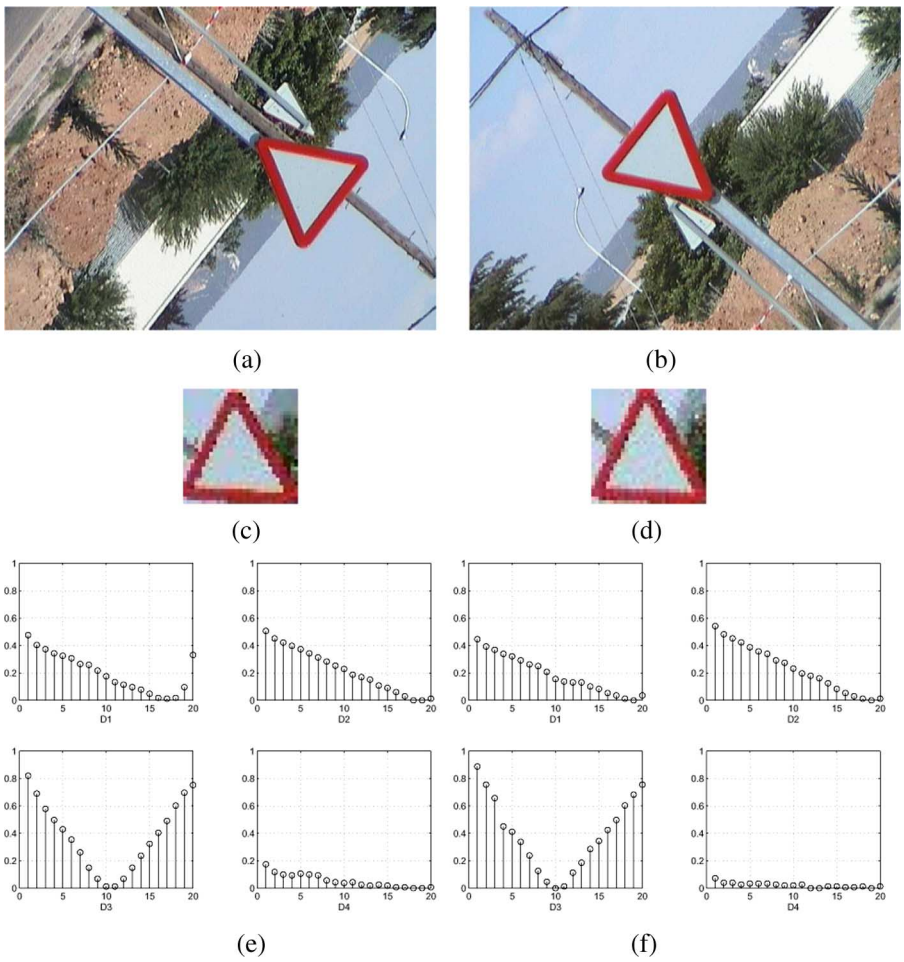


Fig. 12. Rotation 2-D invariance. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

been often used to solve binary classification problems, they can also be applied to regression. For shape classification, we use linear SVMs. In the case of two separable classes, the training data are labeled $\{\mathbf{x}_i, y_i\}$, where $i = 1, \ldots, l$, $y_i \in \{-1, 1\}$, $\mathbf{x}_i \in \{\mathcal{R}^d\}$. In our case, the vectors $\mathbf{x}_i$ are the DtBs, as we will describe later, the values y_i are "1" for one class and "-1" for the others, d is the dimension of the vector, and l is the number of training vectors. If a hyperplane $\{\mathbf{w}, b\}$ separates the two classes, the points that lie on it satisfy $\mathbf{x} \cdot \mathbf{w}^T + b = 0$, where $\mathbf{w}$ is normal to the hyperplane, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin, and $\|\mathbf{w}\|$ is the Euclidean norm of $\mathbf{w}$. In the separable case, the following constraints hold:

$$y_i(\mathbf{x}_i \cdot \mathbf{w}^T + b) - 1 \geq 0 \quad \forall i. \tag{2}$$

The points for which the equality in (2) holds give us the scale factor for $\mathbf{w}$ and b or, equivalently, a constant difference of the unity. These points lie on both hyperplanes $H_1 : \mathbf{x}_i \cdot \mathbf{w}^T + b = 1$ and $H_2 : \mathbf{x}_i \cdot \mathbf{w}^T + b = -1$. Hence, the margin between the two data sets is simply $2/\|\mathbf{w}\|$. The margin between both sets can be maximized by minimizing $\|\mathbf{w}\|^2/2$ subject to the constraints of (2). If we introduce positive Lagrange multipliers (α_i, where $i = 1, \ldots, l$)—one for each of the inequality constraints (equal to the number of training vectors)—the objective now is to minimize Lp given by

$$Lp = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i \left(\mathbf{x}_i \cdot \mathbf{w}^T + b\right) + \sum_{i=1}^{l} \alpha_i. \tag{3}$$

Once the optimization is completed, we simply determine on which side of the hyperplane a given test vector $\mathbf{x}$ lies. That is, to classify it to one class ("1") or to the other ("-1"), the decision function is given by

$$f(\mathbf{x}) = \operatorname{sgn}(\mathbf{x} \cdot \mathbf{w}^T + b). \tag{4}$$

When data sets are nonseparable, we can introduce further cost by introducing positive slack variables ξ_i, $i = 1, \ldots, l$ in the constraint, i.e.,

$$y_i \left(\mathbf{x}_i \cdot \mathbf{w}^T + b\right) - 1 \geq \xi_i \quad \forall i. \tag{5}$$

A natural way to assign extra cost for errors is to change the objective function to be minimized from $\|\mathbf{w}\|^2/2$ to $\|\mathbf{w}\|^2/2 + C(\sum_i \xi_i)^k$, where C is a parameter to be chosen by the user. A larger C corresponds to assignment of higher penalty to errors.

In this paper, we present DtBs as feature vectors for the inputs of the linear SVMs, as introduced in [2]. DtBs are the

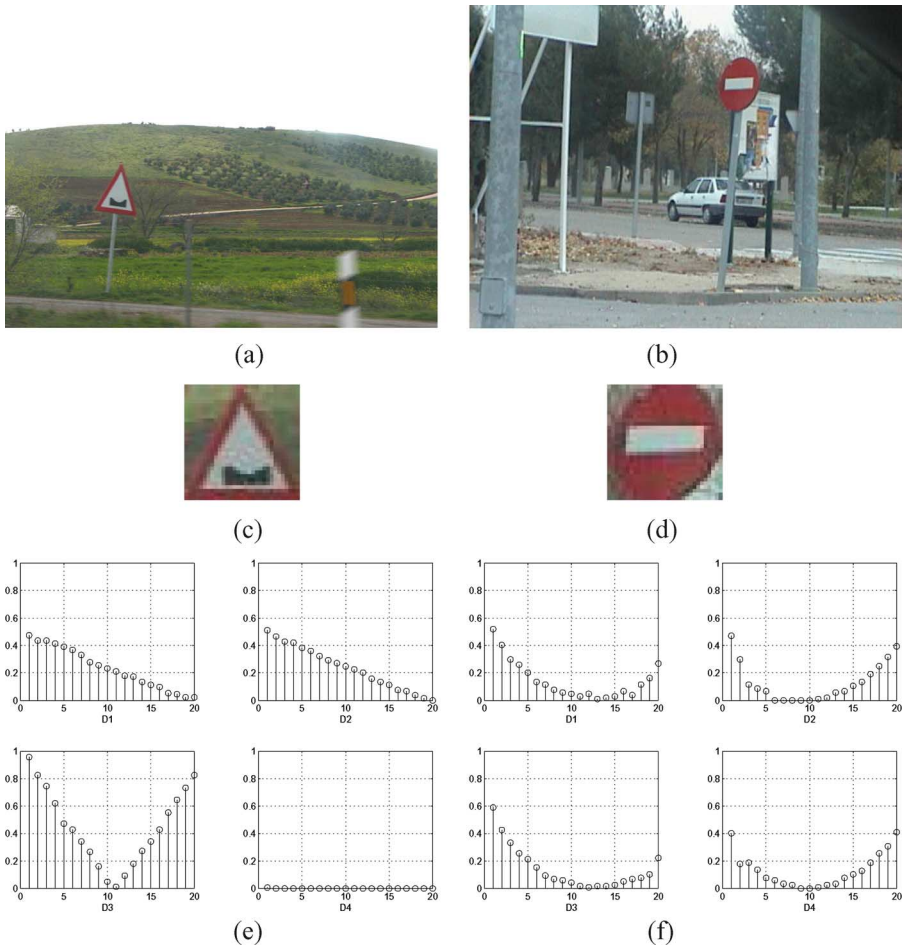


Fig. 13. Rotation 3-D invariance. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

distances from the external edge of the blob to its bounding box. Fig. 9 shows these distances for a triangular shape where $D1$, $D2$, $D3$, and $D4$ are the left, right, upper, and bottom DtBs, respectively.

Four DtB vectors of 20 components are obtained, and they feed specific SVMs depending on the previous color extraction because the segmentation color determines the possible geometric shapes (see Fig. 10), as mentioned previously. We note that the octagonal shape is easily confused with the circular one at medium–high distances, and for this reason, a simple way of characterizing octagonal signs is to consider them as circular and try to identify their message in the recognition stage. That way, an extracted blob by red color feeds four DtB SVMs to classify the shape as a circle ("1") or not ("-1") and another four SVMs to classify the shape as a triangle ("1") or not ("-1"). Thus, four favorable votes are possible for each shape. A majority voting method has been applied in order to get the classification with a threshold; therefore, if the total number of votes is lower than this value, the analyzed blob is rejected as a noisy shape. In case of a tie, linear SVM outputs of favorable classification are used to decide which is the candidate shape.

The proposed method is invariant to translation, rotation, and scale. First, it is invariant to translation because it does not matter where the candidate blob is. Thus, in Fig. 11, the vectors of two signs that are placed in very different positions present high similarity. Second, the detection process is invariant to rotation because, before obtaining DtB vectors, the most external pixels of each blob are detected to determine the original orientation, and after this, all blobs are oriented in a reference position. In conclusion, samples of DtB vectors show a similar evolution for each geometric shape. Figs. 12 and 13 show the vectors for different rotations. Finally, the method is invariant to scale due to the normalization of the DtB vectors to the bounding-box dimensions. Fig. 14 shows the vectors for the same sign that is observed at different distances as we approach it.

Moreover, since four feature vectors are obtained to characterize every blob, this method is quite robust to occlusions. Fig. 15 shows the results of a correct shape classification of signs under partial occlusions. The robustness of the system against occlusions will be analyzed later in Section IV. Results showing these properties were reported in [2].

C. Recognition

Once the candidate blobs are classified into a shape class, the recognition process is initiated. Recognition is implemented by SVMs with Gaussian kernels. For the training process of SVMs, we used the library LIBSVMS [26].

Linear SVMs have been briefly described in Section III-B. However, in many cases, the data cannot be separated by a

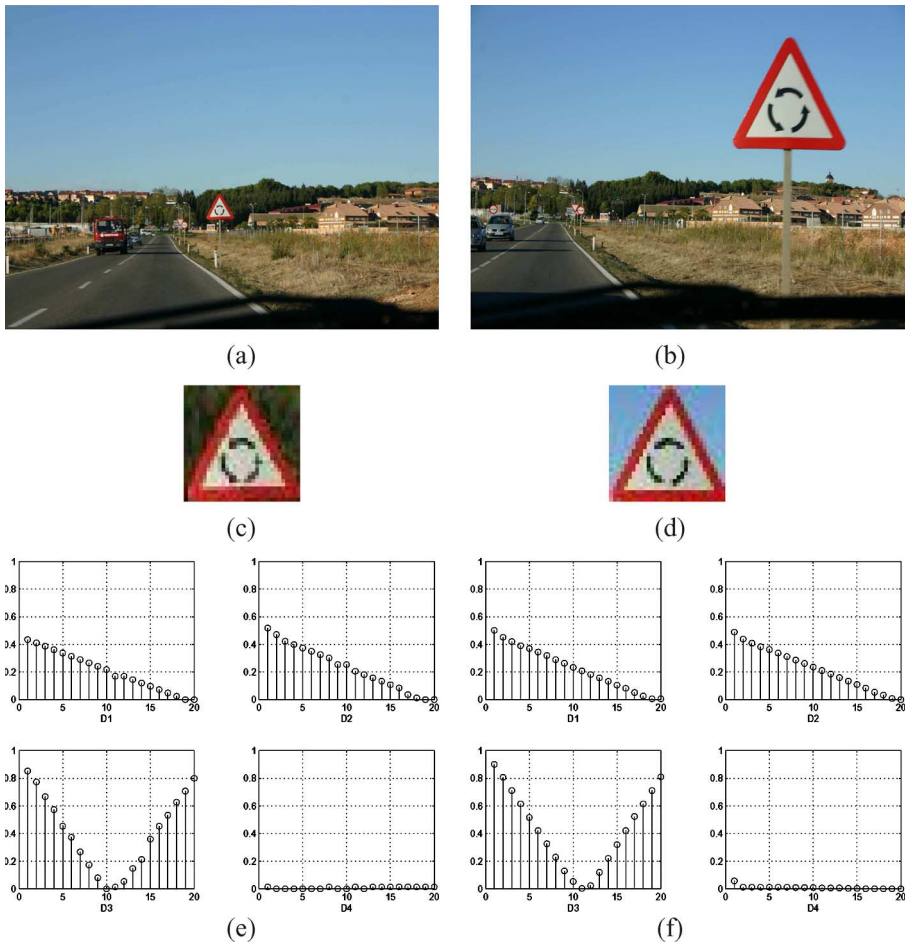


Fig. 14. Scale invariance. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

linear function. A solution is to map the input data into a different space $\Phi(\mathbf{x})$. Due to the fact that the training data are used through a dot product, if there was a "kernel function," so that we satisfy $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$, we can avoid computing $\Phi(\mathbf{x})$ explicitly and use the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$.

In this paper, we have used a Gaussian kernel as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \tag{6}$$

and the decision function for a new input vector is

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b \right) \tag{7}$$

where N_s is the number of support vectors, and $\mathbf{s}_i$ are the support vectors. In this case, the sum cannot be reduced to a single dot product.

The recognition stage input, in our case, is a block of 31 $\times$ 31 pixels in grayscale image for every candidate blob; therefore, the interior of the bounding box is normalized to these dimensions. In order to reduce the feature vectors, only those pixels that must be part of the sign (pixel of interest, PoI) are used. For instance, for a circular sign, only pixels that are inside the inscribed circle, which belong to the normalized bounding box, are computed in the recognition module. Fig. 16 shows the PoI for two signs whose shapes are circular and triangular.

Different one-versus-all SVMs classifiers with a Gaussian kernel are used, so that the system can recognize every sign. Both the training and test are done according to the color and shape of each candidate region; thus, every candidate blob is only compared to those signs that have the same color and shape as the blob to reduce the complexity of the problem.

The amount of training samples per class varies between 20 and 100. We use an average of 50 training patterns for each class, but only some of them define the decision hyperplane as support vectors. Figs. 17 and 18 show the support vectors that define the decision region for a "No overtake" traffic sign when objects are extracted by achromatic segmentation. Of course, the training set includes samples of noisy objects that could be confused with traffic signs by the recognition module. To search for the decision region, all feature vectors of a specific class are grouped together against all vectors corresponding to the rest of classes (including here noisy objects), following the one-versus-all classification algorithm previously mentioned.

In Table II, we present the optimum values for regularization parameters in SVMs C and g, where C is the cost parameter for the slack constraints in (5), and g is the inverse of $2\sigma^2$ in the kernel function in (6). An exhaustive search was performed in order to find the values where the total number of errors in the training process was the lowest. In the test phase, we

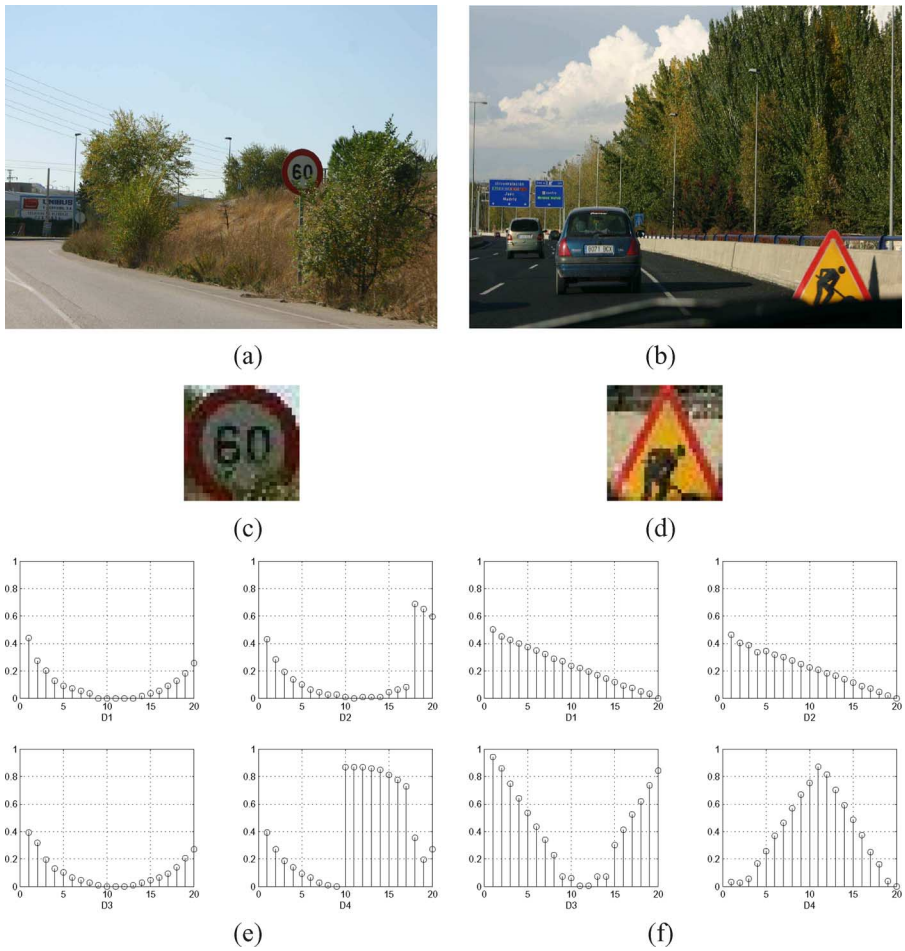


Fig. 15. Occlusions. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

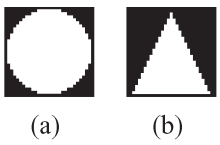


Fig. 16. PoIs in white. (a) Circular and (b) triangular.

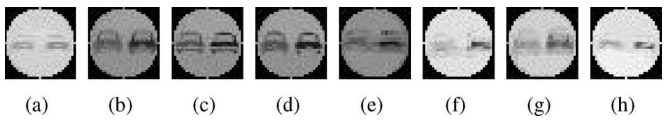


Fig. 17. Positive support vectors for the "No overtake" traffic sign by achromatic segmentation.

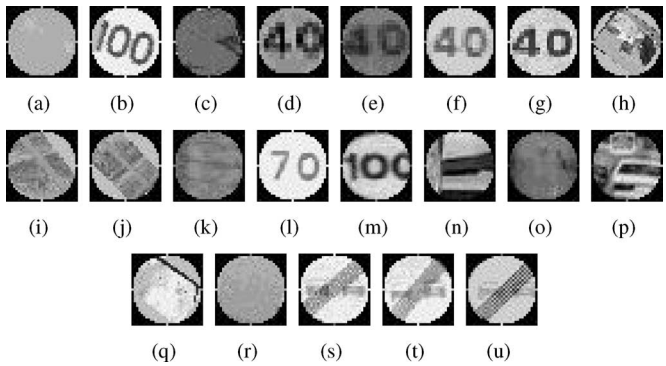


Fig. 18. Negative support vectors for the "No overtake" traffic sign by achromatic segmentation.

assign each BoI the traffic-sign class for which the system obtains the highest value when evaluating the decision function. The threshold values for discarding a blob as noise have been fixed at zero for the decision functions of all SVMs. However, this value can be modified in order to change the false alarm probability and the lost probability.

In the traffic-sign database, there is a set of triangular signs whose pictograms present a high level of similarity at low resolution (see Fig. 19). As the recognition module of our system works with blobs that are normalized to 31 $\times$ 31 pixels, it is not easy to discern between them. This is why, in this paper, as a first approximation, we have reorganized these signs within a unique training set, which will be represented by the sign that is shown in Fig. 19(a).

IV. Results

In our experiments, test sequences have been recorded with a video camcorder (Canon MVX30i) fixed onto the front windshield of a vehicle while driving at usual speed. The video sequences were first converted into ".bmp" images using the

TABLE II
OPTIMUM VALUES FOR PARAMETERS OF ONE-VERSUS-ALL SVMs EMPLOYED IN THE RECOGNITION STEP

Color Segmentation	Shape	C	g
Red	Circle	32	0.00781
Red	Triangle	32	0.00781
White	Circle	32	0.03125
White	Triangle	32	0.00781
Yellow	Circle	32	0.00781
Yellow	Triangle	32	0.00781

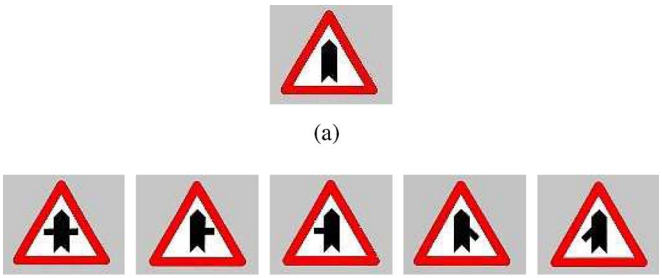


Fig. 19. Set of similar signs. (a) Representative sign of the set. (b) Junction with minor road ahead. (c) Minor road from the right ahead. (d) Minor road from the left ahead. (e) Minor road joins from the left ahead. (f) Minor road joins from the left ahead.

TABLE III
SUMMARY OF RESULTS

Number of sequence	1	2	3	4	5
Number of images	749	1774	860	995	798
Number of traffic signs	21	21	20	25	17
Detections of traffic signs	218	237	227	285	127
Noisy candidate blobs	601	985	728	622	434
False alarms	0	3	4	8	7
Confused recognition	4	4	4	2	7

software DVGrab 1.7.[1] The size of each image is 720 × 576 pixels, and the time between two successive captured frames is 0.2 s.

Currently, in our research, we are particularly concerned with automatic inventory control of traffic signs. For this reason, sequences are run as batch processes. The code to implement the whole algorithm was C, and we achieve a mean processing time of 1.77 s per frame on a 2.2-GHz Pentium 4-M, where the frame dimensions are 720 × 576 pixels.

Several video sequences with the results are available at http://roadanalysis.uah.es/page_files/publications.html. All sequences that are shown have been captured driving at usual speed over a stretch of approximately 4 km and during both the day and at night. In addition, these images cover a wide range of weather conditions: sunny, cloudy, and rainy weather. Sequences 1, 2, and 3 correspond to a sunny lighting day, sequence 4 corresponds to a rainy day, and, finally, sequence 5 was captured at night.

By inspecting the obtained results, we can say that all signs have been correctly detected in each of the five sequences at least twice. Table III summarizes the results that were generated in the processing of video sequences: 1) the number of frames to analyze in each sequence; 2) the number of traffic signs that appear in each sequence; 3) the number of detections of traffic signs that have been identified correctly (obviously, each sign is detected multiple times in the sequence); 4) noisy blobs that are detected by color and shape and have been rejected by the recognition module; 5) false positives at the output of the system; and 6) confused recognition cases. The situations of confused recognition can be attributed to long distances from the sign to the camera or to a poor lighting, in the case of night visibility, as the illumination depends on the vehicle's headlamp. If we consider future tracking performance, we can establish the criterion that a candidate traffic sign is dismissed if it appears in only one frame of the sequence. Thus, each traffic sign will be correctly detected if it is identified at least in two frames of the sequence. An important conclusion from the results is that false alarms do not appear in the same sequence several times, and so they should be rejected by the tracking algorithm. In this respect, Fig. 20 illustrates some examples of correct recognition process although the traffic sign is not identified in all frames of the sequence. In each analyzed image, if there are any BoI, both the shape that is assigned to each BoI and the specific recognized traffic sign are represented. Since, as we explained in Section III-A, a sign can be processed as a sum of two contributions (the rim or outer area and the inner area), the same sign can be detected and recognized once or twice. The shape is always drawn in white color over the bounding box's coordinates, and the identified sign is represented by a synthetic template in the original image. To make the presentation clearly visible, both representations are done independently for every segmentation color.

Figs. 21 and 22 show two subsequences of the experimental results of the sequence S_1, where only eight frames of the route are shown in every case. In Fig. 21, the circular sign is classified and recognized successfully from Fig. 21(d)–(h) by the red color and from Fig. 21(f)–(h) by achromatic segmentation. Note that one or two white contours are drawn in each frame as a result of the shape classification. Thus, the external outline corresponds to segmentation by the red color, and the inside contour corresponds to the achromatic segmentation.

Fig. 23 illustrates that our system also works when the signs are not placed perpendicular to the movement of the vehicle (3-D rotations). Although almost all traffic signs are only affected by affine distortion, our system can detect signs that also present perspective distortion. Thus, even if a sign appears oriented with no ideal position, it can be detected. In addition to the mentioned 3-D rotation, the invariance to translation is supported by the fact that for traffic signs, we search for the whole image, as shown in Figs. 21 and 22. Furthermore, in the same figures, it is illustrated how the system is generally able to recognize objects with so many different scales as standard traffic signs.

Fig. 24 shows an example of road-sign detection and recognition at night. Comparing Figs. 21 and 24, we can observe that the results are similar during both day and night because the material of the road signs is highly reflective at night. Fig. 25 illustrates that our system also works when arrays of two or more traffic signs exist in the image. Here, the system

[1]DVGrab is a software digital video for Linux, and it can be downloaded from http://kinov.org.

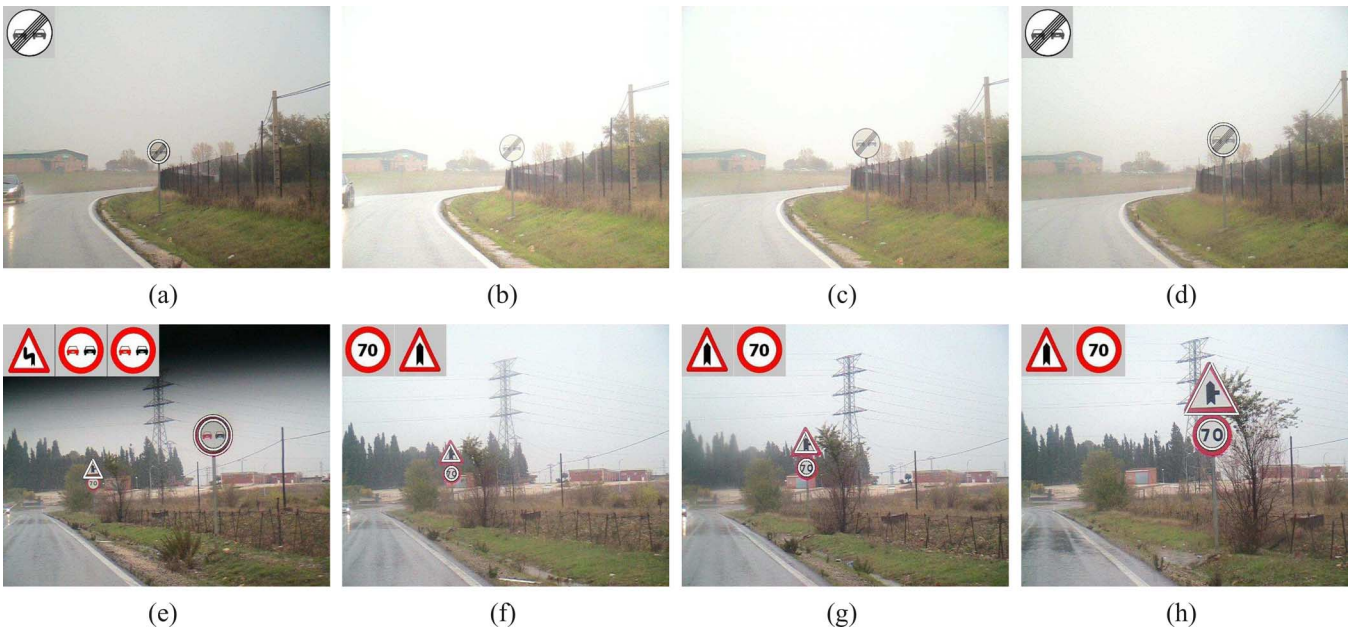


Fig. 20. Examples of recognition process including incorrectly recognized frames. (a)–(d) Recognition with misdetection in (b) and (c). (e)–(h) Recognition with confused classification in (e).

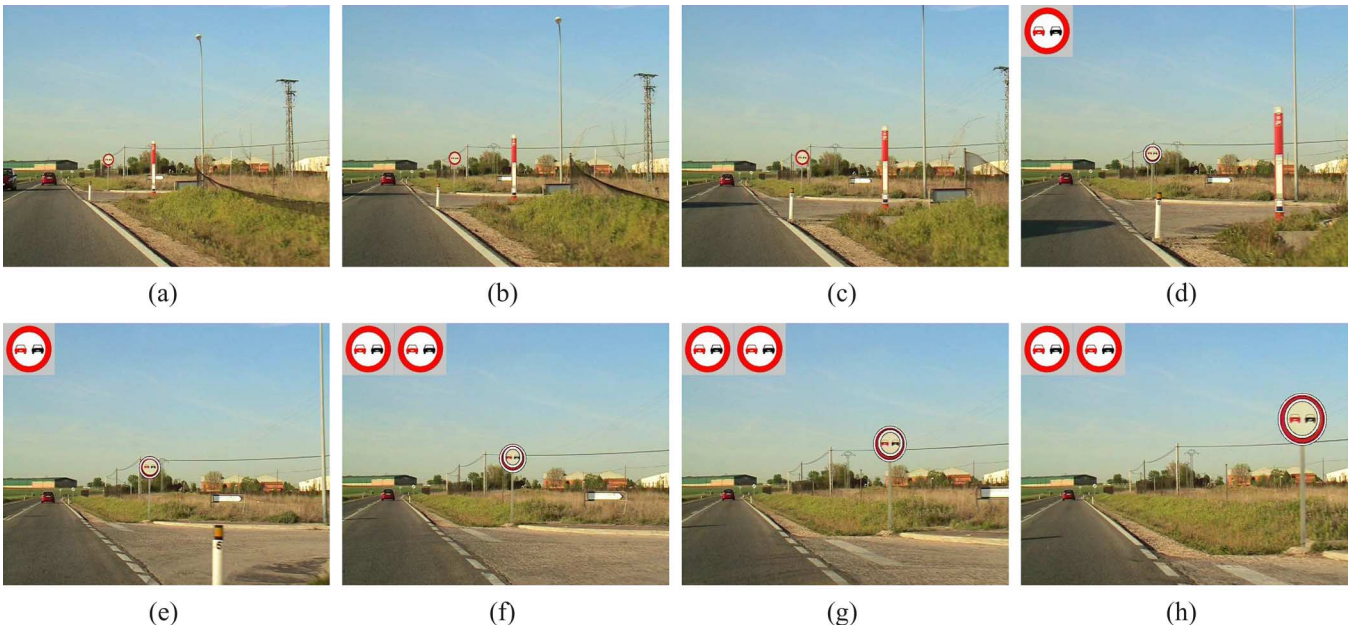


Fig. 21. Experimental results with eight frames for a circular road sign. The size of each image is 720 × 576 pixels, and the time between two successive images is 0.2 s.

isolates the different signs and works by recognizing each sign independently.

The main task in our work is to evaluate the signaling of the road. Hence, in order to reduce the false alarm probability, we have fixed a criterion for discarding small blobs whose dimensions are under 31 × 31 pixels. Obviously, as we approach the signs, the segmentation of the sign becomes better, and the output values of recognizing SVMs are greater than that under normal conditions.

In order to test the behavior of our system to occlusions, a set of signs were manually manipulated to provoke occlusions of different levels in different directions with a circular and a triangular traffic sign. We introduced occlusions with a synthetic circular mask over original images whose color components were assigned with random values between 0 and 255. To test the area that is covered in occlusions, we considered three different sizes. Consequently, the diameters of the occlusion masks were fixed at one half, one third, and one fourth of the larger dimension of the corresponding bounding box (see Fig. 26). Then, eight orientations were analyzed for every mask with increases of 45°, i.e., the same mask was shifted and in that way was placed in these different positions, as illustrated

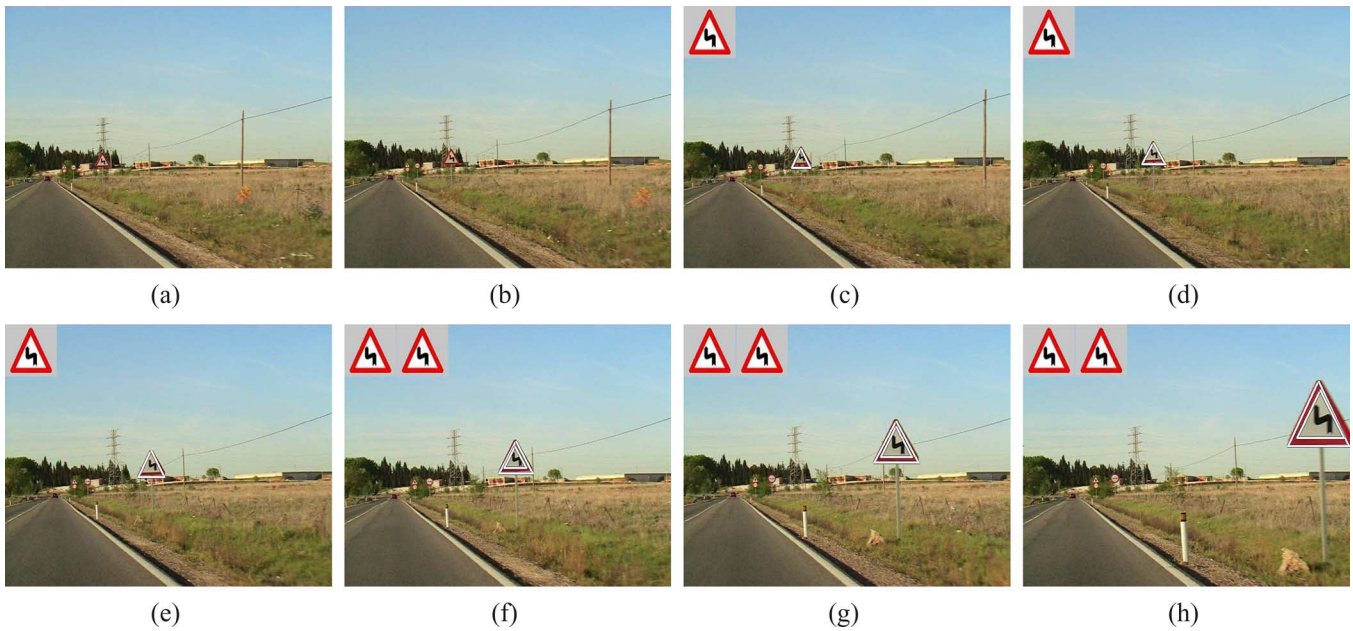


Fig. 22. Experimental results with eight frames for a triangular road sign.

Fig. 23. Experimental results with 3-D rotation.

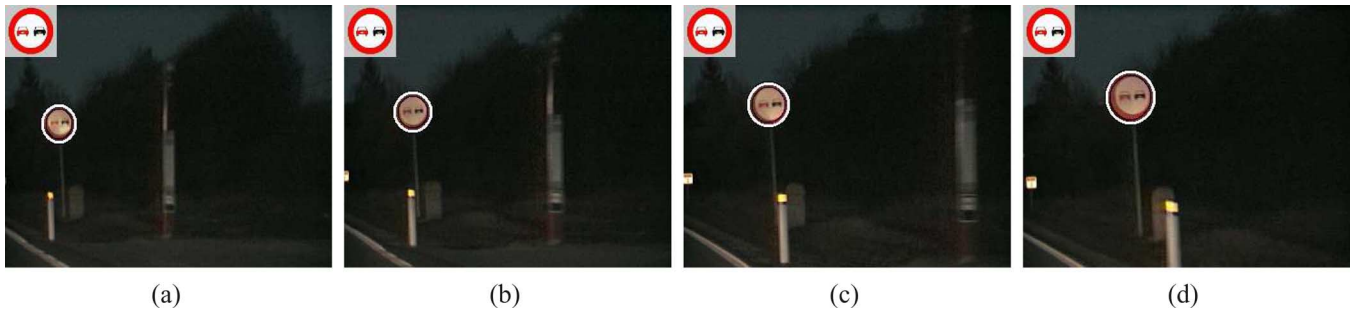


Fig. 24. Experimental results at night.

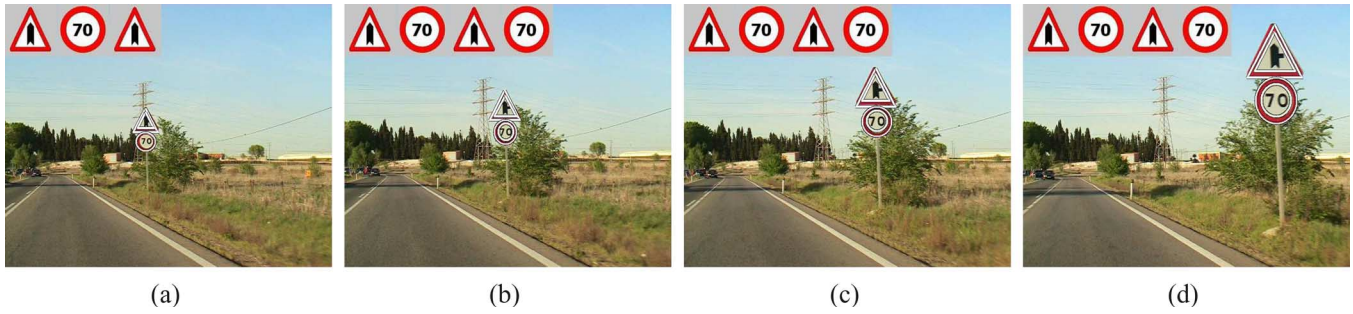


Fig. 25. Experimental results with array road signs.

(a) (b) (c)

Fig. 26. Masks of occlusion. Three sizes are used to test the robustness of the system against occlusions with different areas covered. (a)–(c) Masks whose diameters are one half, one third, and one fourth of the major dimension of the bounding box, respectively.

(a) (b) (c)

Fig. 27. Displacements followed by a mask for the first orientation.

in Fig. 27. The recognition success probabilities are 93.24%, 67.85%, and 44.90% for the small, medium-sized, and large masks, respectively.

By inspecting the results, we can conclude that the worst results in the recognition stage with occlusions are obtained when the occlusion mask is placed in the middle of the pictogram's inner area and, obviously, when the size of the mask is increased.

V. Conclusion

This paper describes a complete method to detect and recognize traffic signs from a video sequence, taking into consideration all the existing difficulties regarding object recognition in outdoor environments. Thus, our system can be useful for the maintenance of traffic signs.

Two main modules have been developed based on the capability of SVMs as a novel technique in pattern recognition: a first module for shape classification based on linear SVMs and a second one that was developed with Gaussian kernels for recognition of the inner area. Although a tracking method has not been developed yet, we have considered a candidate sign as valid in a sequence if it is detected and recognized in at least two frames of a sequence. Otherwise, it is considered to be a false alarm.

Experimental results indicate that our system is accurate because it allows us to detect different geometric shapes, i.e., circular and octagonal, and triangular and rectangular, and works correctly in difficult situations, for example, when array signs appear on the scene or when images are taken under night illumination. Moreover, as we have seen before, the system is invariant to rotations, changes of scale, and different positions. In addition, the algorithm can also detect signs that are partially occluded. However, some improvements remain as tasks for the future.

- We must recognize rectangular route-guidance signs for navigation whose pictograms can present very different icons.
- In Fig. 20(b) and (c), the sign is not detected due to the similarity of the sign and the background. A new segmentation method can be developed based on analysis of color similarity between neighbor pixels and not be strictly oriented to the segmentation of color.
- Since our main application is the maintenance of traffic signs, we hope to extract all the information about the state of traffic signs in future works, such as the level of deformation, angle of rotation, and conservation state.
- Obviously, a real-time implementation is an improvement for future works; in this way, two main issues arise to reduce processing time per frame. The first is to pack recognition information, leading to a reduction of the necessary feature space dimensions. The second is a possible reduction of the frames' dimensions to be analyzed.

Acknowledgment

The authors would like to thank the anonymous reviewers and the editors for their many helpful suggestions, which have greatly improved the presentation of this paper.

References

[1] L. Fletcher, N. Apostoloff, L. Petersson, and A. Zelinsky, "Vision in and out of vehicles," *IEEE Intell. Syst.*, vol. 18, no. 3, pp. 12–17, May/Jun. 2003.

[2] S. Lafuente-Arroyo, P. Gil-Jiménez, R. Maldonado-Bascón, F. López-Ferreras, and S. Maldonado-Bascón, "Traffic sign shape classification evaluation I: SVM using distance to borders," in *Proc. IEEE Intell. Veh. Symp.*, Las Vegas, NV, Jun. 2005, pp. 557–562.

[3] P. Gil-Jiménez, S. Lafuente-Arroyo, H. Gomez-Moreno, F. López-Ferreras, and S. Maldonado-Bascón, "Traffic sign shape classification evaluation II: FFT applied to the signature of blobs," in *Proc. IEEE Intell. Veh. Symp.*, Las Vegas, NV, Jun. 2005, pp. 607–612.

[4] H. Kamada, S. Naoi, and T. Gotoh, "A compact navigation system using image processing and fuzzy control," in *Proc. Southeastcon*, New Orleans, LA, Apr. 1990, vol. 1, pp. 337–342.

[5] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE Trans. Ind. Electron.*, vol. 44, no. 6, pp. 848–859, Dec. 1997.

[6] J. Miura, T. Kanda, and Y. Shirai, "An active vision system for real-time traffic sign recognition," in *Proc. IEEE Intell. Transp. Syst.*, Oct. 2000, pp. 52–57.

[7] P. Arnoul, M. Viala, J. Guerin, and M. Mergy, "Traffic signs localisation for highways inventory from a video camera on board a moving collection van," in *Proc. IEEE Intell. Veh. Symp.*, Tokyo, Japan, Sep. 1996, pp. 141–146.
[8] A. de la Escalera, J. M. Armingol, J. M. Pastor, and F. J. Rodríguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 57–68, Jun. 2004.
[9] C. Y. Fang, C. S. Fuh, P. S. Yen, S. Cherng, and S. W. Chen, "An automatic road sign recognition system based on a computational model of human recognition processing," *Comput. Vis. Image Underst.*, vol. 96, no. 2, pp. 237–268, Aug. 2004.
[10] H. Liu, D. Liu, and J. Xin, "Real-time recognition of road traffic sign in motion image based on genetic algorithm," in *Proc. 1st. Int. Conf. Mach. Learn. Cybern.*, Nov. 2002, pp. 83–86.
[11] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 566–571.
[12] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proc. Int. Conf. IROS*, Sendai, Japan, Sep./Oct. 2004, pp. 70–75.
[13] Y. Aoyagi and T. Asakura, "A study on traffic sign recognition in scene image using genetic algorithms and neural networks," in *Proc. 22nd IEEE Int. Conf. Ind. Electron., Control Instrum.*, Taipei, Taiwan, R.O.C., Aug. 1996, vol. 3, pp. 1838–1843.
[14] C. Fang, S. Chen, and C. Fuh, "Road sign detection and tracking," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1329–1341, Sep. 2003.
[15] A. Farag and A. E. Abdel-Hakim, "Detection, categorization and recognition of road signs for autonomous navigation," in *Proc. Adv. Concepts Intell. Vis. Syst.*, Brussels, Belgium, Aug./Sep. 2004, pp. 125–130.
[16] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th. Int. Conf. Comput. Vis.*, Corfu, Greece, Sep. 1999, pp. 1150–1157.
[17] A. de la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image Vis. Comput.*, vol. 21, no. 3, pp. 247–258, Mar. 2003.
[18] E. Perez and B. Javidi, "Nonlinear distortion-tolerant filters for detection of road signs in background noise," *IEEE Trans. Veh. Technol.*, vol. 51, no. 3, pp. 567–576, May 2002.
[19] S. H. Hsu and C. L. Huang, "Road sign detection and recognition using matching pursuit method," *Image Vis. Comput.*, vol. 19, no. 3, pp. 119–129, 2001.
[20] W. Wu, X. Chen, and J. Yang, "Detection of text on road signs from video," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 4, pp. 378–390, Dec. 2005.
[21] M. Thomson and S. Westland, "Colour-imager characterization by parametric fitting of sensor responses," *Colour Res. Appl.*, vol. 26, no. 6, pp. 442–449, Dec. 2001.
[22] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[23] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[24] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998. [Online]. Available: citeseer.ist.psu.edu/burges98tutorial.html
[25] N. Cristianini and J. Shame-Taylor, *Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[26] C. Chang and C. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Saturnino Maldonado-Bascón (S'98–A'00–M'04) received the Telecommunication Engineering degree from the Universidad Politécnica de Madrid, Madrid, Spain, in 1996 and the Ph.D. degree from the Universidad de Alcalá, Alcalá de Henares, Madrid, in 1999.

Since 1997, he has been a member of the faculty of the Universidad de Alcalá, where he became an Associate Professor with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, in 2001. His research interests are image processing and pattern recognition.

Sergio Lafuente-Arroyo received the B.S. and M.S. degrees from the Universidad de Alcalá, Alcalá de Henares, Madrid, Spain, in 2000 and 2004, respectively, both in telecommunication engineering.

He is currently an Instructor with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, Universidad de Alcalá. His research interests include digital image processing and computer vision.

Pedro Gil-Jiménez received the Telecomunication Engineering degree from the Universidad de Alcalá, Alcalá de Henares, Madrid, Spain, in 2001.

In 1999, he joined the Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, Universidad de Alcalá, where he has been a Lecturer since 2001. His research interests include computer vision, video surveillance systems, and intelligent transportation systems, particularly those involving signal and image processing.

Hilario Gómez-Moreno (M'99) received the B.S. degree in telecommunication engineering from the Universidad de Alcalá, Alcalá de Henares, Madrid, Spain, in 1994 and the M.S. degree in telecommunication engineering from the Universidad Politécnica de Madrid in 1999.

Since 1997, he has been with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Politécnica Superior, Universidad de Alcalá, where he became a Full-Time Lecturer in 2002. His research interests are image processing and support vector machines.

Francisco López-Ferreras received the Telecommunication Engineering degree and the Ph.D. degree from the Universidad Politécnica de Madrid, Madrid, Spain, in 1981 and 1992, respectively.

He is currently with the Universidad de Alcalá, Alcalá de Henares, Madrid, where he has been an Associate Professor with the Departamento de Teoría de la Señal y Comunicaciones since 1998 and is the Dean of the Escuela Politécnica Superior. His research interests include signal processing.