

Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification

Xavier Baró, Sergio Escalera, Jordi Vitrià, Oriol Pujol, and Petia Radeva

Abstract—The high variability of sign appearance in uncontrolled environments has made the detection and classification of road signs a challenging problem in computer vision. In this paper, we introduce a novel approach for the detection and classification of traffic signs. Detection is based on a boosted detectors cascade, trained with a novel evolutionary version of Adaboost, which allows the use of large feature spaces. Classification is defined as a multiclass categorization problem. A battery of classifiers is trained to split classes in an Error-Correcting Output Code (ECOC) framework. We propose an ECOC design through a forest of optimal tree structures that are embedded in the ECOC matrix. The novel system offers high performance and better accuracy than the state-of-the-art strategies and is potentially better in terms of noise, affine deformation, partial occlusions, and reduced illumination.

Index Terms—Dissociated dipoles, ensemble of dichotomizers, Error-Correcting Output Code (ECOC), evolutionary boosting, traffic sign recognition.

I. INTRODUCTION

TRAFFIC sign recognition is being studied for several purposes such as autonomous driving and assisted driving [1], [2]. Recognition of traffic signs allows a driver to be warned about inappropriate actions and potentially dangerous situations. In the mobile mapping framework, traffic-sign-recognition methods are used with other methods to compile road information and measure the position and orientation of different landmarks in movement either on an aerial or a terrestrial platform. One example of this system is given by Madeira *et al.* [3], where a mobile mapping system automatically processes traffic signs. In this work, a recognition accuracy of more than 80% on a reduced set of sign types is obtained. In [4], a vehicle-based vision platform is used to detect road signs, where the main goal is mainly focused on speed signs.

In the literature, we can find two main approaches to solve the problem of road sign recognition: 1) color based and

2) gray-scale based. Color-based sign recognition relies on color to reduce false-positive results in the recognition process [5]–[13], whereas the gray-scale method concentrates on the geometry of the object [14]–[17]. Recent works use the combination of both cues to improve the detection rates. For instance, in [18], a threshold is applied over a hue, saturation, value representation of the image to find regions with a high probability of having a traffic sign. Many background objects can share colors with traffic signs; thus, heuristics over the size and aspect ratio are used to reduce the number of false-alarm regions. Once the regions are normalized to a predefined size, a linear support vector machine (SVM) is used to classify the region in one of the possible shapes (e.g., circle or triangle). The color and shape information is used as a coarse classification, and an SVM with Gaussian kernels is used to perform the fine-classification step. The color information is strongly related to the type of camera, illumination, and sign aging; thus, the use of color information introduces additional difficulties to recognition. In the work of de la Escalera *et al.* [19], these difficulties are addressed using an enhancement step before the use of thresholds on the color values. After applying size heuristics to remove nonsign regions, the authors use a fusion of color information, the gradient, and a distance image to remove regions with low probability of having a traffic sign. Final classification is performed through a neural network. Other recent works are focused on the final classification step. Paclik *et al.* [20] propose a representation of road-sign data based on extending the traditional normalized cross-correlation approach to a similarity based on individual matches in a set of local image regions.

Traffic sign recognition is a straightforward application for object recognition algorithms in which the previous addressing of the category detection (e.g., object location) is often required. In recent years, one of the most accepted and widely used approaches in object detection has been the one proposed by Viola and Jones [21]. Their approach is based on a cascade of detectors, where each detector is an ensemble of boosted classifiers based on the Haar-like features. Lienhart and Maydt [22] presented an extension of the original Haar-like features set, demonstrating that *Adaboost* converges faster and with better results when the features set is large. On the other hand, due to the exhaustive search over the features set, the training time grows with respect to the number of features. This fact makes any approach that tries to extend the feature set unfeasible.

Once an object (i.e., a traffic sign) is located, it should be recognized from a wide set of possible classes using some kind of classification technique. Designing a machine-learning

Manuscript received October 18, 2007; revised April 11, 2008 and October 2, 2008. First published February 2, 2009; current version published February 27, 2009. This work was supported in part by a research grant under Projects TIN2006-15308-C02, FISPI061290, and CONSOLIDER-INGENIO 2010 (CSD2007-00018) and in part by the *Institut Cartogràfic de Catalunya* under the supervision of M. Pla. The Associate Editor for this paper was C. Stiller.

X. Baró is with the Computer Vision Center, Edifici O, Campus Universitat Autònoma de Barcelona, 08193 Barcelona, Spain (e-mail: xbaro@cvc.uab.cat).

S. Escalera, J. Vitrià, O. Pujol, and P. Radeva are with the Computer Vision Center, Edifici O, Campus Universitat Autònoma de Barcelona, 08193 Barcelona, Spain, and also with the Department Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, 08007, Barcelona, Spain (e-mail: sescalera@cvc.uab.cat; jordi@cvc.uab.cat; oriol@cvc.uab.cat; petia@cvc.uab.cat).

Digital Object Identifier 10.1109/TITS.2008.2011702

multiclass technique is a hard task. In this sense, it is common to conceive algorithms to distinguish between just two classes and combine them in some way. Following the multiclass categorization problem, where a set of classifiers should learn, in a natural way, the features shared between categories, the Error-Correcting Output Code (ECOC) technique was proposed with very interesting results [23]. This technique is a very successful multiclass categorization tool due to its ability to share the classifiers knowledge among classes. Recently, the embedding of a tree structure in the ECOC framework has shown to be highly accurate with a very small number of binary classifiers [24]. However, the ECOC design is still an open issue.

The goal of this paper is twofold. First, during detection, to discriminate a set of traffic signs from the background, we propose a novel binary classifier through an evolutionary version of Adaboost, which avoids the limitations of the original algorithm in terms of the dimensionality of the feature space. Second, to deal with the multiclass categorization problem to distinguish among a large set of classes, a multiclass learning technique is proposed. The approach is based on embedding a forest of optimal tree structures in the ECOC framework, allowing for sharing of features (e.g., tree nodes, base classifiers, and dichotomies) among classes in a very robust way. Finally, we develop a real traffic-sign-recognition system for a mobile mapping process [25], in which we validate the robustness of our approach. We show that the present strategy obtains high accuracy and outperforms the state-of-the-art strategies and is robust against a high variability of sign appearance.

This paper is organized as follows. Section II presents the novel detection and multiclass categorization approaches. Section III shows the architecture of the real traffic sign recognition system and integration details. Section IV shows the experimental results, and Section V concludes this paper.

II. METHODOLOGY

In this section, we present the new evolutionary boosting strategy and the Forest-ECOC (F-ECOC) technique to deal with the object detection and classification stages, respectively.

A. Detection

Detection takes an image as input and gives, as the output, the regions that contain the candidate object. Our detector is inspired by detector presented by Viola and Jones [21]. We consider the use of the *attentional cascade* concept, boosting as the feature selection strategy, and the representation of the image in terms of the integral image. With our approach, we solve the limitation of the exhaustive search of the boosting formulation in [21]. Moreover, we present a solution to the restriction of the boosting-process computation when the feature set size is large. We propose an evolutionary vision of boosting, which not only drastically reduces the training time but also allows the use of huge feature sets. This way, we propose the use of the dissociated dipoles—a more general type of features than the Haar-like features [21]—which can also be calculated using the integral image. In addition, we introduce another

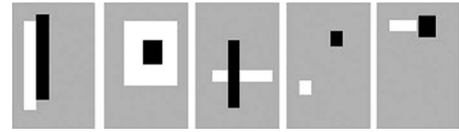


Fig. 1. Dissociated dipoles. The black region corresponds to the *inhibitory dipole*, whereas the white region corresponds to the *excitatory dipole* [27].

variation in the *Weak Classifier*, changing the decision rule from the threshold value that was used in the original schema to its ordinal definition, where only the sign is considered. This representation has been demonstrated to be more robust in the case of noise and illumination changes [26].

1) *Detection Architecture*: Working with unbalanced problems like object detection, each time we analyze an image, the system must discard a large number of negative regions, whereas just a few or any negative region corresponds to the object that we are looking for. The attentional cascaded architecture allows for discarding easy nonobject regions at low computational cost, whereas more complex regions are deeply analyzed. One attentional cascade is composed of a set of classifiers or stages, where the input to each classifier corresponds to regions that are classified as object-regions by the previous stages. The regions classified as object-regions by the last stage are the output of the detector.

Although any learning method can be used to learn the cascade, the usefulness of Adaboost has widely been demonstrated. Before each stage training, a new set of samples is built using the positive samples and the false positives of the previous stages of the cascade. Then, a classifier is learned to achieve a given minimum hit and maximum false-alarm rates. This process is repeated until the desired number of stages or the target false-alarm rate are reached.

2) *Dissociated Dipoles*: In [22], Lienhart and Maydt show that the accuracy of the detector increases with the number of available features. Therefore, in this paper, we use the *dissociated dipoles*—a more general type of features than the Haar-like features—and, thus, deal with a larger feature set.

The *dissociated dipoles* or *sticks* have been presented by Balas and Sinha [27], and these sticks are features that are composed of a pair of rectangular regions, named the *excitatory dipole* and the *inhibitory dipole*, respectively (see Fig. 1). The mean value of all the pixels in the *inhibitory dipole* is subtracted from the mean value of the *excitatory-dipole* pixels. As in the case of the Haar-like features, the integral image is used to calculate the sum of the pixels inside the rectangular regions. Similar features are also used in other recent works, for example, [28], where weighted dipoles with a fixed position are used to describe objects.

In computational terms, the use of dissociated dipoles means increasing from the approximately 600.000 features in Lienhart's approach of to more than 2^{30} features in a training window size of 30×30 pixels, making the classical approach computationally unfeasible. To deal with this limitation, we define an evolutionary approach of Adaboost.

3) *Evolutionary Adaboost*: Boosting is a powerful learning technique that allows for combining the performance of many simple classification functions or *Weak Classifiers* to produce a

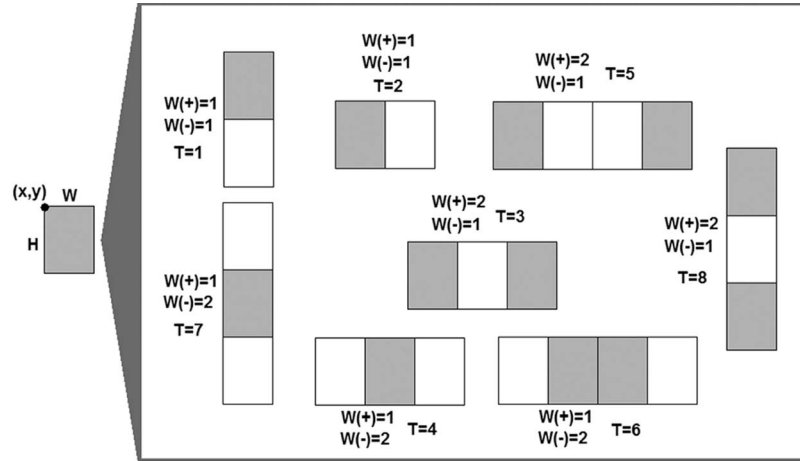


Fig. 2. Haar-like features definition from a given region and the type value.

Strong Classifier [29]. At each round of learning, the examples are reweighted to emphasize the examples that were incorrectly classified by the previous *Weak Classifier*. The final *Strong Classifier* is a decision stump, which is composed of a weighted combination of *Weak Classifiers*, followed by a threshold [30].

In the classical boosting approach, an exhaustive search is used to find the best *Weak Classifier*. Therefore, with an enormous features set, this approach becomes computationally unfeasible.

Our *evolutionary weak learner* minimizes the weighted error function ε of the Adaboost scheme as

$$\varepsilon = \sum_{i:h(x_i) \neq y_i} w_i \quad (1)$$

where $X = \{(x_i, y_i) | i = 1 : m\}$ are the pairs sample-label that compound the training set, $W = \{w_1, w_2, \dots, w_m\}$ is the Adaboost weights distribution over the training set, and $h(x_i)$ corresponds to the label that the hypothesis h predicted for the training object x_i . Although, in the rest of this section, we use the original *Discrete Adaboost* algorithm for simplicity, our approach can be applied to any existing variant of Adaboost.

Weak Learner can be considered an optimization problem, where we need to find the parameters of the Weak Classifier that minimize the error function ε . This function is defined; thus, it seems logical to consider it as a nonderivative function that is full of discontinuities. Therefore, the classical approaches based on gradient descend cannot be applied on this problem. One alternative solution is the use of an evolutionary approach. The most well known evolutionary strategy is the genetic algorithm, which does a search over the spaces of solutions using the three basic concepts of the Darwin's theory: 1) mutation, 2) crossover, and 3) natural selection.

4) *Weak Classifier*: When we work with evolutionary algorithms, we should define two basic elements: 1) the individuals and 2) the evaluation function. Each individual must represent a point in the space of solutions, and the evaluation function measures the function value at the point that the individual represents. We define the evaluation function as $\mathcal{F}(I) \mapsto \mathbb{R}$,

$I \in \mathbb{R}^d$, where I is an individual. In this case, the function that we are optimizing is the classification error over a given labeled data set X , using a weights distribution W over the data. Then, the function \mathcal{F} corresponds to ε [see (1)], where the individual I defines the hypothesis h . Combining both equations, we can write the evaluation function as

$$\mathcal{F}(I, W, X) = \left(\sum_{i:h(I, x_i) \neq y_i} w_i \right)$$

where $h(I, x) \mapsto \{-1, +1\}$. (2)

Note that the function h depends not only on a certain data instance but also on the individual. At this point, we formulated the evaluation function in terms of the *Weak Learner*. The following step determines which parameters are required to define h or, in other terms, to decide the dimension of I . One important consideration in choosing the parameters is to evaluate the relevance of each of them in contrast to the introduced complexity.

The Viola and Jones definition of the *Weak Learner* consists of a feature and a threshold value. The feature can be parameterized by the upper left position and the size of one of the regions and their type. The size, weight, and position of the other regions that conform to the Haar-like feature are fixed by the feature type (see Fig. 2). Given a feature, the threshold value must be learned using the training samples, applying an exhaustive search over all the possible threshold values to find the one that minimizes the error. Therefore, for the Viola approach, a *Weak Learner* can be defined as $h(I, Thr, x_i) \mapsto \{-1, +1\}$, where $I = (R_x, R_y, R_w, R_h, T)$, Thr is the threshold value, R_x, R_y, R_w , and R_h correspond to the upper left corner (x, y) , the width, and the height of one of the regions, and T is the type of the Haar-like feature (see Fig. 2). Note that the threshold value is not included in the individual I , because this value is learned once the other parameters have been fixed.

In the case of the Dissociated Dipoles, the regions have no constraints; thus, both regions are independently learned. Using the same reasoning, we define a Weak Classifier based on

the Dissociated Dipoles as $h(I, Thr, x_i) \mapsto \{-1, +1\}$, where $I = (Re_x, Re_y, Re_w, Re_h, Ri_x, Ri_y, Ri_w, Ri_h)$, with Re being the excitatory dipole, and the type parameter T is changed by the parameters of the inhibitory dipole Ri . This representation can be extended by including extra parameters such as the weights of the regions, which are necessary to represent some types of Haar-like features.

In the previous approaches, the difference between both regions depends on illuminance conditions. Thus, we need to define a normalization criterion to make the representation invariant to those conditions. In [21], a method called *fast lighting correction* is used to deal with the illuminance variations. In addition, the Haar-like approach uses the sum of the pixels to represent the region, which introduces a scale dependence that must be corrected to deal with the multiscale detection. The dissociated dipoles are not affected by scale, because the mean value is used instead of the sum.

The aforementioned formulations are quantitative comparisons between the regions, considering not only which region has a higher value but the difference between these values as well. We can simplify by only using qualitative comparisons; therefore, the calculus of the threshold value becomes unnecessary, because we use only the sign of the difference. This approach has two main advantages: 1) The illuminance normalization is unnecessary, and 2) removing the threshold learning process reduces the evaluation time. We can rewrite the previous formulations as

$$h(I, x_i) \mapsto \{-1, +1\} \quad (3)$$

where

$$I = (R_x, R_y, R_w, R_h, T)$$

$$I = (Re_x, Re_y, Re_w, Re_h, Ri_x, Ri_y, Ri_w, Ri_h).$$

5) *Learning Algorithm*: The final approach is summarized in the Evolutionary Adaboost algorithm, as shown in Algorithm 1. This algorithm is used to learn all the stages of the detection cascade using a set $\langle (x_1, y_1), \dots, (x_m, y_m) \rangle$ of samples classified as positive samples in the previous stages of the cascade. This algorithm iteratively uses a genetic algorithm to minimize the weighted error and to instantiate the parameters of a new *Weak classifier* that is added to the final ensemble.

Algorithm 1: Evolutionary Discrete Adaboost.

Given: $(x_1, y_1), \dots, (x_m, y_m)$
 where $x_i \in X, y_i \in Y = \{-1, +1\}$
 Initialize $W_1(i) = 1/m$
for $t = 1, \dots, T$ **do**
 Use a genetic algorithm to minimize

$$\epsilon_t = Pr_{i \sim W_t} [h_t(x_i) \neq y_i].$$

The given solution is taken as the hypothesis h_t .
 Get the weak hypothesis $h_t : X \mapsto \{-1, +1\}$ with error ϵ_t .

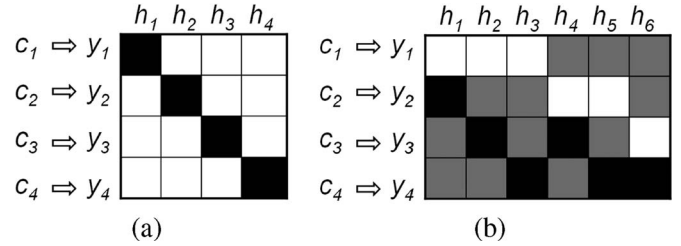


Fig. 3. Four-class ECOC designs. (a) One-versus-all ECOC codification. (b) One-versus-one ECOC codification (White: 1. Black: -1. Grey: 0).

Choose $\alpha_t = (1/2) \ln(1 - \epsilon_t/\epsilon_t)$

Update

$$W_{t+1}(i) = \frac{W_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t}, & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t}, & \text{if } h_t(x_i) \neq y_i \end{cases} \\ = \frac{W_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that W_{t+1} will be a distribution).

end for

Output the final hypothesis:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right).$$

B. Classification: F-ECOC

Once we have located an object, we need to categorize among a large set of classes. Although various systems of multiple classifiers were proposed, most of them use similar constituent classifiers, which are often called base classifiers (i.e., dichotomies). In this sense, ECOC represents a classification technique that allows a successful combination of base classifiers to address the multiclass problem [23], [31].

1) *ECOC*: The design of ECOC is based on a coding and a decoding strategy, where coding aims at assigning a codeword¹ to each of the N_c classes (up to N_c codewords), and decoding aims at assigning a class label to a new test codeword. Arranging the codewords as rows of a matrix, we define the “coding matrix” M , where $M \in \{-1, 1\}^{N_c \times n}$, with n being the code length. From the point of view of learning, the matrix M represents n binary learning problems (i.e., dichotomies), with each corresponding to a column of the ECOC matrix M . Each dichotomy defines a subpartition of classes, coded by $\{+1, -1\}$ according to their class membership. In Fig. 3(a) the codification for a four-class problem using the one-versus-all coding strategy is shown. The white and black regions correspond to +1 and -1 valued positions, respectively. Thus, in Fig. 3(a), the dichotomy h_i is trained to discriminate class c_i from the rest of the classes. If we use a larger set of symbols for coding $M \in \{-1, 0, 1\}^{N_c \times n}$, some entries in the matrix M can be zero, which indicates that a particular class is not considered for a given dichotomy. In Fig. 3(b), the codification for a four-class problem that uses the one-versus-one coding strategy is

¹A codeword is a sequence of bits that represents a class.

shown. The gray regions correspond to the zero value (i.e., nonconsidered classes for the classifiers). In this strategy, all possible pairs of classes are split. For example, dichotomy h_1 classifies class c_1 versus class c_2 .

As a result of the outputs of the n binary classifiers, at the decoding step, a code is obtained for each data point in the test set. This code is compared with the base codewords of each class that was defined in the coding matrix M , and the data point is assigned to the class with the “closest” codeword. The common distances for decoding are the Hamming and the Euclidean distances [32].

2) *F-ECOC*: Most of the current discrete coding strategies are predesigned problem-independent codewords (e.g., one versus all [33] and one versus one [34]). In the work of Pujol *et al.* [24], a method for embedding tree structures in the ECOC framework is proposed. Beginning on the root containing all classes, the nodes associated with the best partition in terms of the mutual information are found, and the process is repeated until the sets with a single class are obtained.

Taking the previous work as a baseline, we propose to use multiple trees embedding, forming an F-ECOC. We build an optimal tree—the one with the highest classification score at each node—and several suboptimal trees—the ones closer to the optimal one under certain conditions. Let us keep, at each iteration, the best k partitions of the set of classes. If the best partition is used to construct the current ECOC tree, the rest of the partitions form the roots of $k - 1$ trees. We iteratively repeat this process until all nodes from the trees are decomposed into one class. Given a base classifier, the suboptimal tree candidates are designed to have the maximum classification score at each node without repeating the previous subpartitions of classes. In the case of generating T first optimal trees, we can create an ensemble of trees by embedding them in the ECOC matrix, as shown in Algorithm 2.

Algorithm 2: Training algorithm for F-ECOC.

Given N_c classes: c_1, \dots, c_{N_c} and T trees to be embedded

$\Omega_0 \leftarrow \emptyset$

$i \leftarrow 1$

for $t = 1, \dots, T$ **do**

 Initialize the tree root with the set $N_i = \{c_1, \dots, c_{N_c}\}$

 Generate the best tree at iteration t :

for each node N_i **do**

 Train the best partition of its set of classes $\{P_1 P_2\} | N_i = P_1 \cup P_2, N_i \notin \Omega_{t-1}$ using a classifier h_i so that the training error is minimal

 According to the partition obtained at each node, codify each column of the matrix M as:

$$M(r, i) = \begin{cases} 0, & \text{if } c_r \notin N_i \\ +1, & \text{if } c_r \in P_1 \\ -1, & \text{if } c_r \in P_2 \end{cases}$$

 where r is the index of the corresponding class c_r

$\Omega_t \leftarrow \Omega_{t-1} \cup N_i$

$i \leftarrow i + 1$

end for

end for

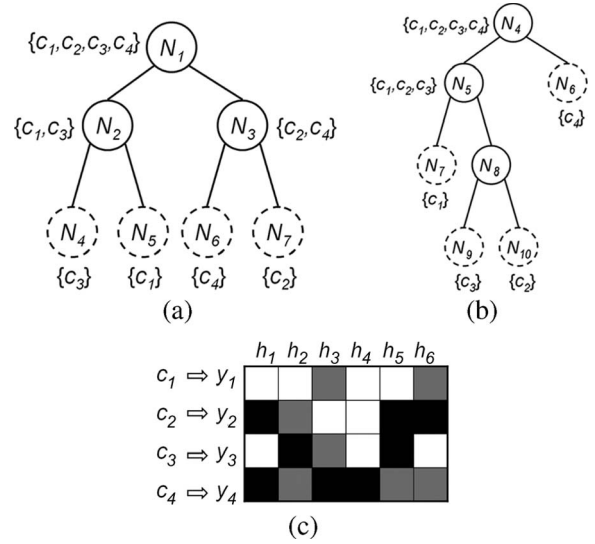


Fig. 4. Four-class optimal trees and the F-ECOC matrix. (a) First optimal tree for a four-class problem. (b) Second optimal tree for the same problem. (c) F-ECOC matrix M for the problem, where h_1, h_2 , and h_3 correspond to classifiers of N_1, N_2 , and N_3 from the first tree, and h_4, h_5 , and h_6 to N_4, N_5 , and N_8 from the second tree.

The proposed technique provides a suboptimal solution because of the combination of robust classifiers obtained from a greedy search using the classification score. One of the main advantages of the proposed technique is that the trees share their information among classes in the ECOC matrix M . It is done at the decoding step by jointly, instead of separately, considering all the coded positions of a class. It is easy to see that each tree structure of N_c classes introduces $N_c - 1$ classifiers, which is far from the $N_c \cdot (N_c - 1)/2$ dichotomies required for the one-versus-one coding strategy.

One example of two optimal trees and the F-ECOC matrix for a toy problem is shown in Fig. 4. Fig. 4(a) and (b) show two examples of optimal trees. The second optimal tree is constructed based on the following optimal subpartitions of classes. This way, for the first initial set of classes $\{c_1, c_2, c_3, c_4\}$, the two optimal trees include the best subpartitions of classes in terms of the classification score, which, in the example, corresponds to c_1, c_3 versus c_2, c_4 for the first tree and c_1, c_2, c_3 versus c_4 for the second tree, respectively. Fig. 4(c) shows the embedding of trees into the F-ECOC matrix M . Note that the column h_3 corresponds to the node N_3 , and the following dichotomies correspond to the nodes of the second tree. Classes that do not belong to the subpartitions of classes are set to zero. On the other hand, classes that belong to each partition are set to $+1$ and -1 values, defining the subset of classes involved on each classifier.

Recent studies on the decoding steps have shown that the zero symbol introduces decoding errors in the traditional decoding distances [35]. To deal with this problem and to increase the performance of the F-ECOC coding design, we propose the Attenuated Euclidean decoding strategy, which is defined as

$d_j = \sqrt{\sum_{i=1}^n |y_i^j| (x_i - y_i^j)^2}$, where d_j is the distance to row j , n is the number of dichotomies, x_i is the response of the classifier h_i over the test sample, and y_i^j is the value of the coding matrix M at the i th row and the j th column, respectively. We introduce the factor $|y_i^j|$ to avoid the error that the zero symbol introduces.

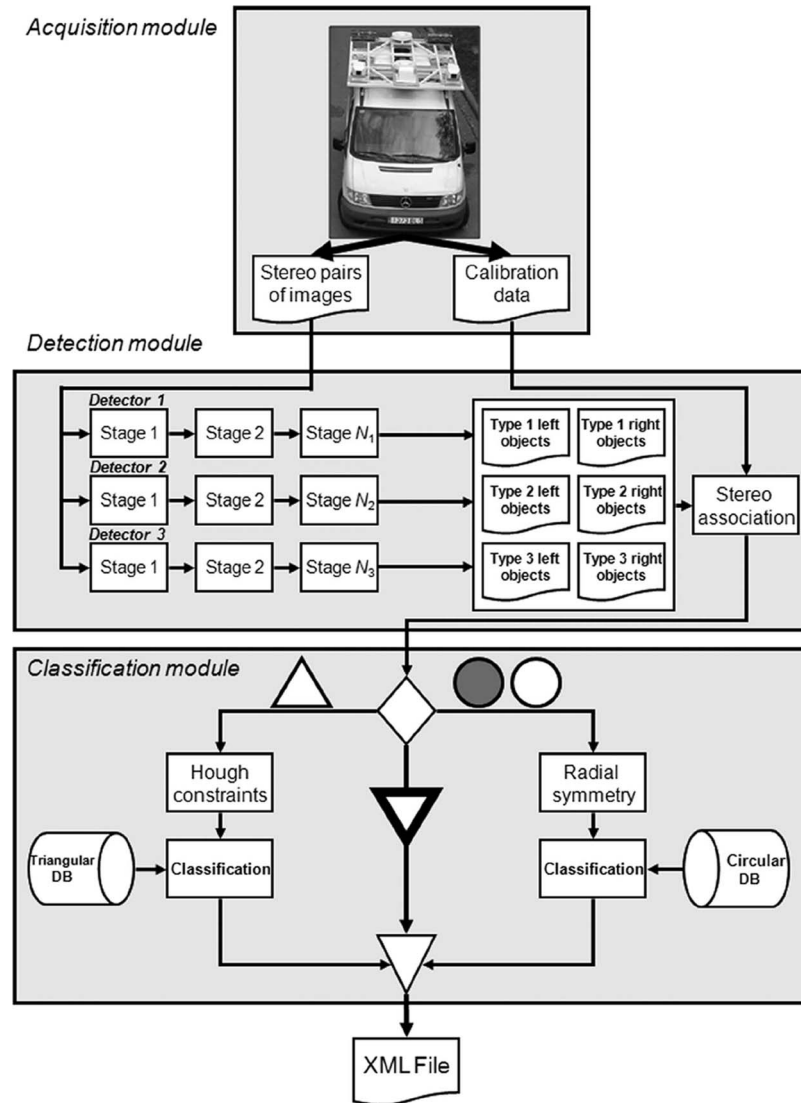


Fig. 5. Scheme of the whole traffic sign recognition system.

III. TRAFFIC-SIGN-RECOGNITION SYSTEM

This section presents the details of the system scheme in Fig. 5. We explain the relationship between each of the aforementioned methods and their integration in a real traffic-sign-recognition system.

A. Acquisition Module

The mobile mapping system has a stereo pair of calibrated cameras, which are synchronized with a Global Positioning System/(GPS/INS) system. Therefore, the result of the acquisition step is a set of stereo pairs of images with their position and orientation information. This information allows the use of epipolar geometry to change from one camera to the other one and to obtain the real position in world coordinates of a point. To detect all signs that appear in the input image independent of their size and position, we scan the image using windows at different scales. All these windows are used as the detector input. In this sense, the detector can be defined as a classification function $h : X \mapsto \{-1, 1\}$, where X is a window from the input im-

age. The result of detection is a set of windows that corresponds to the valid objects, i.e., all the windows X , where $h(X) = 1$.

B. Detection Module

All the data generated in the acquisition process is given to the detector. We organize the trained detectors as an attentional cascade [21]. The *attentional cascade* is a degenerated decision tree where, at each stage, a detector is trained to detect almost all objects of interest while rejecting a certain fraction of the nonsign patterns. Because of the huge number of different traffic signs types, we group them using a similarity criterion, and we train a different cascade for each group. Each window in the input images is analyzed by all the cascades, and the detected objects from each cascade are given as output from the detector. Each cascade detects only a certain category of signs; thus, the output objects have the first classification information.

1) *Stereo Association*: We work with a stereo system; thus, all signs appear on both cameras at each frame (except in case of occlusions or if one of the signs is out of the field of view). This redundant information is used to improve the detection

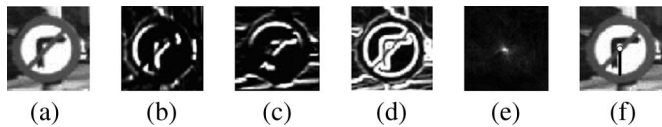


Fig. 6. (a) Input image. (b) X-derivative. (c) Y-derivative. (d) Image gradient. (e) Accumulator of orientations. (f) Center and radius of the sign.

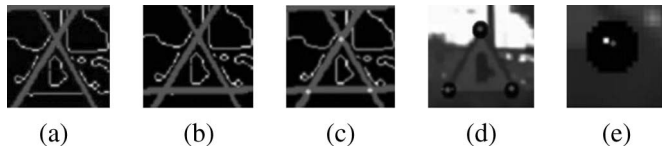


Fig. 7. (a) Detected lines. (b) Corrected line. (c) Intersections. (d) Corner region. (e) Corner found.

ratio. Using the epipolar geometry, given an instance of a sign in one of the sources, we estimate the region where it must appear in the other source. Once we have a reduced search window in the other source, we apply a similarity criterion based on a normalized correlation. The point with the highest similarity value gives us the position of the target object. This information is used to link the object of a source with its stereo object to recover it. Using this information, we only lose the objects that have been lost in both cameras. Using the calibration data, the position and orientation information of the GPS/INS system, and the coordinates of an object in each camera, we compute the object position in world coordinates.

C. Classification Module

Using the Evolutionary Adaboost, a region of interest that contains a sign is determined. Depending on the type of the detected sign, a different model fitting is applied before classification, looking for affine transformations that perform the spatial normalization of the object.

1) *Model Fitting*: Because of the few changes on the point of view of the captured signs, we apply the fast radial symmetry [36] for the circular signs, which offers high robustness to image noise. As shown in Fig. 6, the fast radial symmetry provides an approximation to the center and the radius of the circular sign.

On the other hand, for the case of triangular signs, the method that allows a successful model fitting is based on the Hough transform [37]. Nevertheless, we need to consider additional constraints to obtain the three representative border lines of a triangular traffic sign. Each line has associated a position in relation to the others. In Fig. 7(a), a false horizontal line is shown. This line does not fulfill the expected spatial constraints of the object; thus, we iterate the Hough procedure to detect the next representative line in the allowed range of degrees. The corrected image is shown in Fig. 7(b). Once we have the three detected lines, we calculate their intersection, as shown in Fig. 7(c). To assure that the lines are the expected ones, we complement the procedure by looking for a corner at the circular region of each intersection surroundings [as shown in Fig. 7(d) and (e)] $S = \{(x_i, y_i) | \exists p < ((x - x_i)^2 + (y - y_i)^2 - r^2)\} | i \in [1, \dots, 3]$, where S is the set of valid intersection points, and p corresponds to a corner point to be located

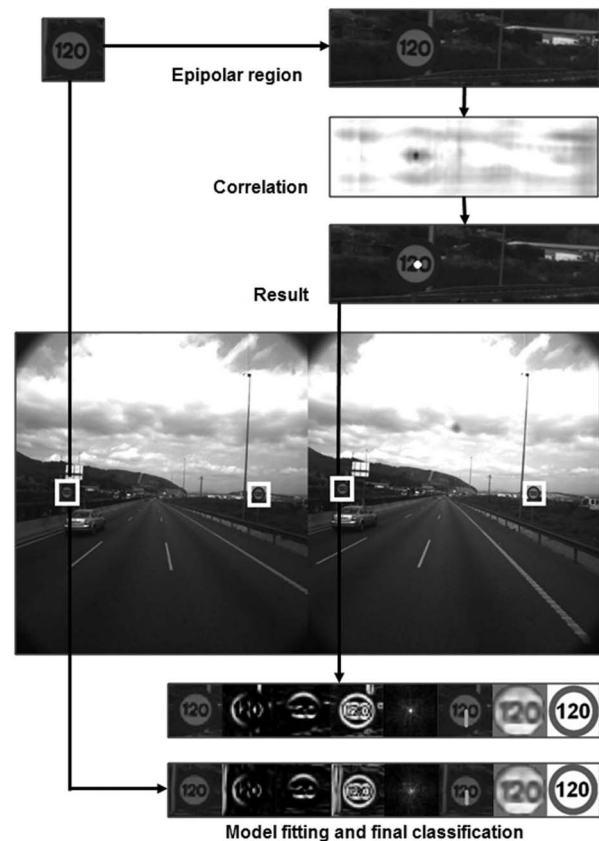


Fig. 8. Example of the system execution. Image acquisition, detection, and classification.

in a neighborhood of the intersection point. A corner should be determined at each lines intersection; thus, we apply a corner detector at the surroundings of the triangular vertices to increase the confidence of determining a sign.

2) *Spatial Normalization*: Once the sign model is fitted using the previous methods, the next step is the spatial normalization of the object before classification. There are four steps.

- 1) Transform the image to make the recognition invariant to small affine deformations.
- 2) Resize the object to the signs database size.
- 3) Filter using the Weickert anisotropic filter [38].
- 4) Mask the image to exclude the background pixels at the classification step.

To prevent the effects of illumination changes, the histogram equalization improves the image contrast and yields a uniform histogram.

3) *F-ECOC*: Once the signs are extracted, they are classified through the F-ECOC strategy. The optimal trees consider the best subpartitions of classes to obtain robust classifiers based on the gray-level pixel values of the object.

D. System Outputs

At the end of the system, an XML file is generated, which contains the position, size, and class of each of the detected traffic signs. This information is used to obtain the real-world position of each detected and recognized sign. One execution example of the whole process is shown in Fig. 8.

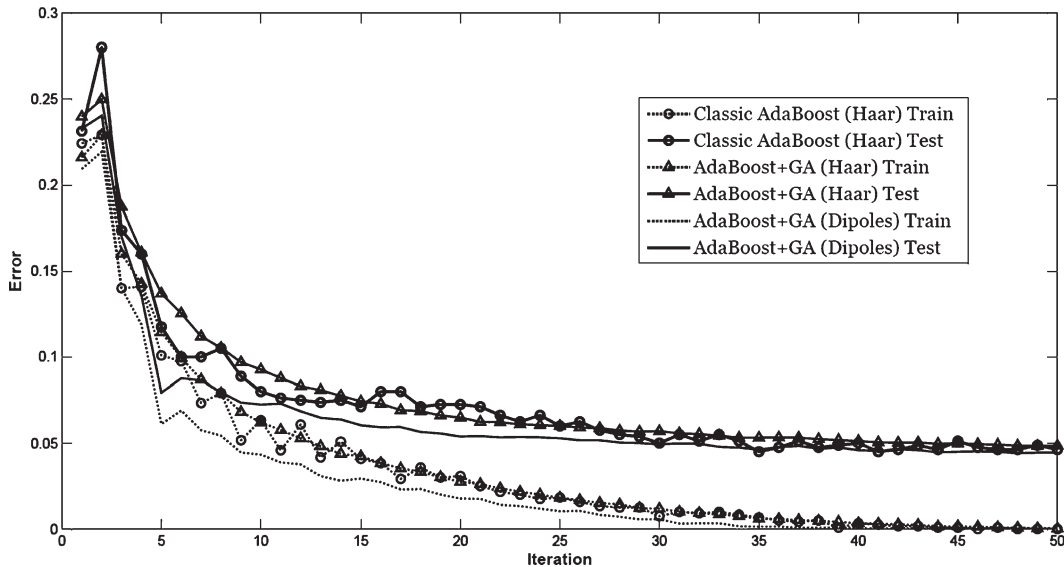


Fig. 9. Error evolution using the classic Adaboost approach and the genetic Weak Learner.

IV. RESULTS

The different experiments of this section are focused on separately (i.e., detection and classification) evaluating each individual part of the framework and performing the whole system evaluation. The validation of the methodology is carried out using real images from the mobile mapping system Geomobil [25]. This system captures georeferenced stereo pairs of images, which are the input of the recognition system. For these experiments, the system was configured to acquire one frame each 10 m or when the orientation variation is greater than 60° . These parameters are hardly tested to assure a correct acquisition in road scenarios, which means that all the interesting objects appear, at least, in two or three stereo pairs of images with a minimum size of 30×30 pixels resolution to be processed.

To assure a high diversity of road conditions, we selected three recording sessions, each one carried out different days and with different weather conditions. It represents a total of 9,510 stereo-pair road images. To avoid using different images that contain the same traffic sign in the training and test sets, instead of using random sequences, we divide the sessions into four subsequences of similar number of frames, without sharing signs. The first and third parts are used for training, and the second and fourth parts are for testing. The reason for using a larger test set is because there are many frames that do not contain objects, and it is interesting to extract the false-alarm ratio in normal conditions, assuring to test the system under different illumination conditions, road types, and traffic intensities.

A. Detection Results

1) *Evolutionary Strategy*: To perform all the tests with the evolutionary Adaboost approach, we use a Genetic Algorithm with a population size of 100 individuals, Gaussian-based mutation probability (i.e., the Gaussian is centered at zero, with a variance of the half of the variable range, decreasing the variance along the generations), and scattered crossover strategy, with a crossover fraction of 0.8. When we use a genetic

algorithm instead of an exhaustive search, different initializations of the algorithm with the same training data give rise to different weak classifiers. The dissociated dipoles cannot be learned by the classical approach; thus, in this experiment, we use the Haar-like features. A one-stage detector is learned using fixed training and test sets, comparing the error evolution for both strategies, and the variance in the evolutionary approach over different runs.

We run the learning process over the same training and test sets for 50 times, using 50 iterations of the evolutionary Adaboost. In the case of the classic Adaboost, because the *Weak Learner* does an exhaustive search over the features, at each round, the selected features are the same. In the case of the Genetic Algorithm, we calculate the mean error value over all the rounds for each iteration.

In Fig. 9, the train and test mean error values at each iteration are shown. Note that both methods converge with the same number of iterations. To analyze the error variability, in Fig. 10, we show the mean and standard deviation for the error at each iteration. The confidence interval shows that the variance is very closed. Therefore, although the evolutionary Adaboost has a random component, the goodness of the given solution is similar.

2) *Detector Performance*: To evaluate the detector performance, we train a cascade of detectors using the evolutionary method with ordinal dissociated dipoles. In Fig. 11, we show the most relevant features that the evolutionary method selected at the first stage of the cascade. Note that only a few of them correspond to Haar-like features.

Due to the different appearance frequency of each type of sign and the high intraclass variability, we trained a detection cascade for each group of similar signs. In Table I, we show the groups of signs and the number of positive samples that were used to train each cascade. The number of negative samples on the train process is automatically selected at each stage, with a proportion of 3 : 1 (i.e., three negative examples for each positive example). Most of the captured images is from main roads, and consequently, some types of signs do not appear enough

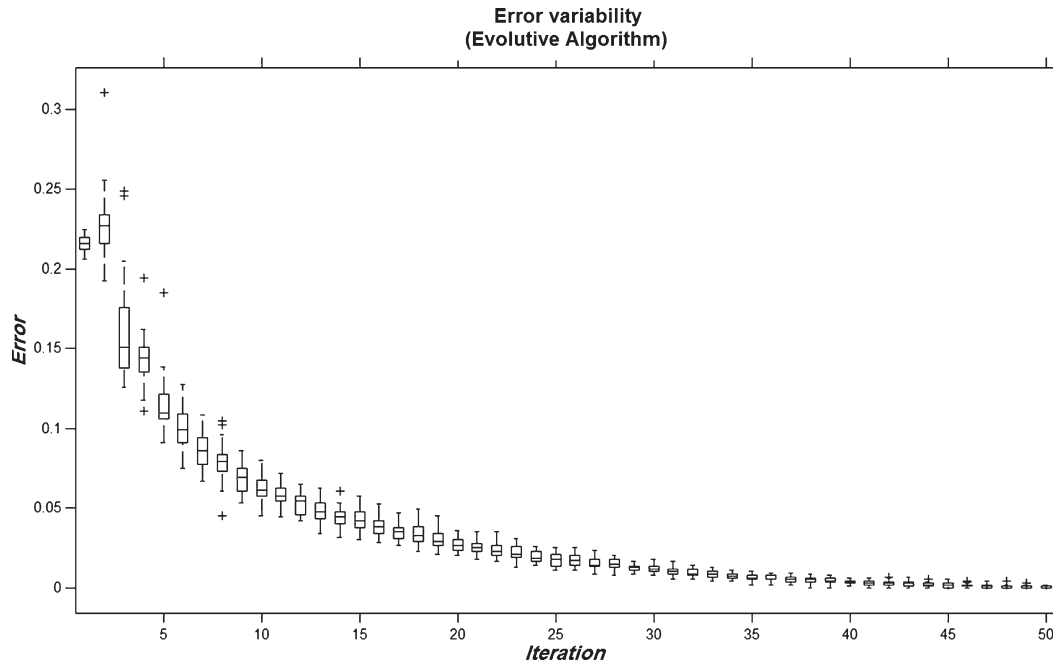


Fig. 10. Genetic approach. Error variability on the training process.



Fig. 11. Selected dipoles obtained from the danger signs.

TABLE I
NUMBER OF POSITIVE SAMPLES USED TO TRAIN THE CASCADE FOR EACH CONSIDERED SIGN

Sign	Danger	Yield	Command	Prohibition
#Samples	545	425	256	993

times to train a detector. Therefore, we only trained the four detectors shown in Table I.

The results are analyzed using two configurations: 1) using the stereo association to take advantage of the stereo information and 2) considering each stereo pair of images as two independent images. For each configuration, the obtained results with and without using sequential information are extracted. When the sequential information is used, different instances of the same real traffic sign are considered as the same object. In the case of not using this information, each instance is considered to be an independent object. In Fig. 12, we show the hit ratio of the detector trained for each type of sign. In general, we can see that the accuracy of the detectors depends on the variability of sign appearance and the size of the training set. The first and the third columns correspond to the results that

consider each appearance of a traffic sign as a different sign. In addition, the second and the fourth columns only take into account the real traffic signs, considering that a sign is detected if we can detect it in one or more frames where it appears. The first two columns do not take into account stereo redundancy, whereas the two last columns do.

Another measure for evaluating the performance of the system is the false-alarm rate. We work with a mobile mapping system; thus, an important point is which percentage of the detected objects corresponds to traffic signs. Therefore, our false-alarm value is referred to the detected signs instead of the number of analyzed windows, which is on the order of 5 million per stereo pair. Nevertheless, the number of false positives with respect to the number of stereo-pair images has been included to analyze the results easier. Both false-alarm rates for each type of signs are detailed in Table II. Some samples of detected objects and false alarms are shown in Fig. 13. It is shown that the system can detect the signs in very extreme lighting conditions. In the false-positive images, it is shown that, frequently, other road elements look similar to traffic signs.

B. Classification Results

Four types of experiments are performed to evaluate the classification scheme:

- 1) traffic sign classification using different state-of-the-art classifiers;
- 2) tree-embedding analysis using F-ECOC;
- 3) public University of California, Irvine (UCI) Machine Learning Repository classification;
- 4) model-fitting classification.

First, we comment the generation of the database to train the classifiers and to perform the experiments.

1) *Classification Database:* The database used to train the classifiers was designed using the regions of interest obtained

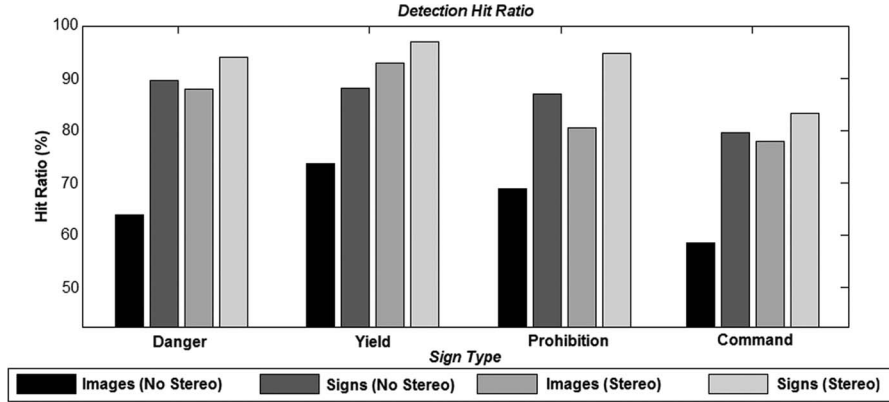


Fig. 12. Hit ratio for each sign type by using dissociated dipoles.

TABLE II
FALSE-ALARM RATES FOR EACH SIGN TYPE

Sign	Danger	Yield	Command	Prohibition
FA/Sign	2.140	4.549	8.551	0.696
FA/Frame	0,045	0,056	0,073	0,019



Fig. 13. Some samples of detected objects and false positives.

from the detection step and the aforementioned model-fitting methods. We defined three groups of classes using the most common types of signs. The considered classes are shown in Fig. 14. Speed signs need special attention. These types of signs are less discriminative, with some of them being only differentiated by a few pixels. With this type of signs, it is better to work on binary images to avoid the errors that can be accumulated because of the gray levels of the signs. For the 12 classes of circular signs and 12 classes of triangular signs, we have 750 training images in both cases. For the seven speed classes, we use 500 training samples. Finally, the resolution of each database is 35×35 pixels for the circular group, 44×39 pixels for the triangular group, and 41×41 pixels for the speed group, respectively.

2) *State-of-the-Art Comparison:* To evaluate the F-ECOC performance, we compare it with state-of-the-art classifiers. The details for each strategy are given as follows:

- 1) Three-Euclidean distance nearest neighbors (K-NN);
- 2) Tangent Distance (TD) [39] with an invariant tangent vector with respect to translation, rotation, and scaling;
- 3) Principal Components Analysis (99.98%), followed by 3-nearest neighbors (PCA K-NN) [40];
- 4) Fisher Linear Discriminant Analysis (FLDA) with a previous 99.98% PCA [40];

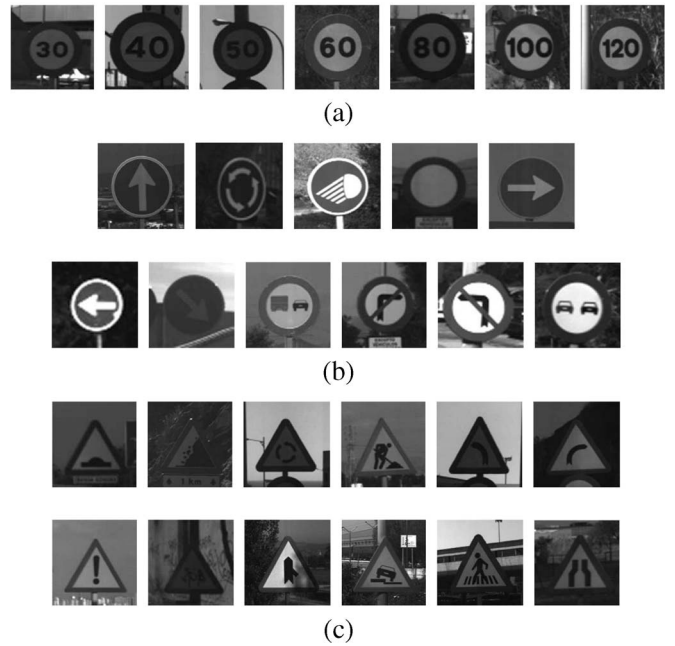


Fig. 14. Set of classes in the classification module. (a) Speed classes. (b) Circular classes. (c) Triangular classes.

TABLE III
CHARACTERISTICS OF THE DATABASES USED FOR CLASSIFICATION. THE TRAINING AND TEST EXAMPLES ARE EQUALLY DISTRIBUTED AMONG THE GROUP CLASSES

Dataset	#Training examples	#Test examples	#Features	#Classes
Circular	750	200	1225	12
Speed	500	200	1681	7
Triangular	750	200	1716	12

- 5) SVM with projection kernel Radial Basis Function and the parameter $\gamma = 1$ (SVM) [41];
- 6) Gentle Adaboost with decision stumps that use the Haar-like features (BR) [22], [42];
- 7) multiclass Joint Boosting (JB) with decision stumps [43];
- 8) Gentle Adaboost [44];
- 9) Sampling with FLDA (BS);
- 10) statistical Gentle Naive Boosting (NB) with decision stumps [42];
- 11) our F-ECOC with 3-embedded optimal trees.

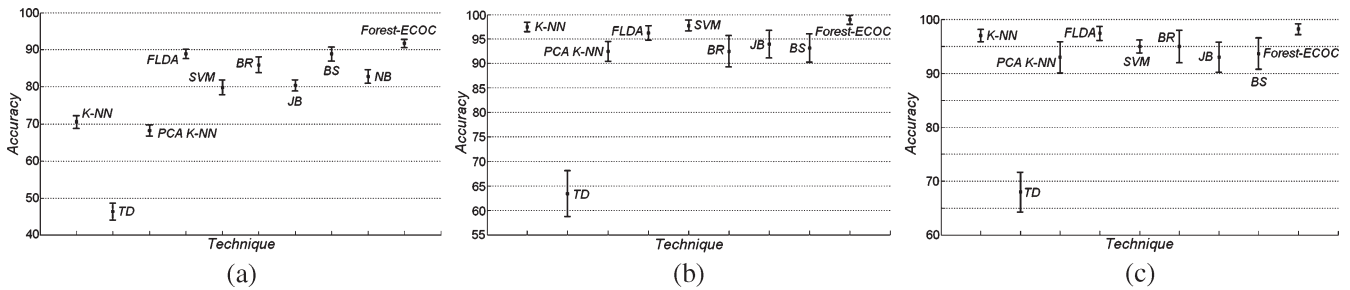


Fig. 15. Classification results. (a) Speed problem. (b) Circular problem. (c) Triangular problem.

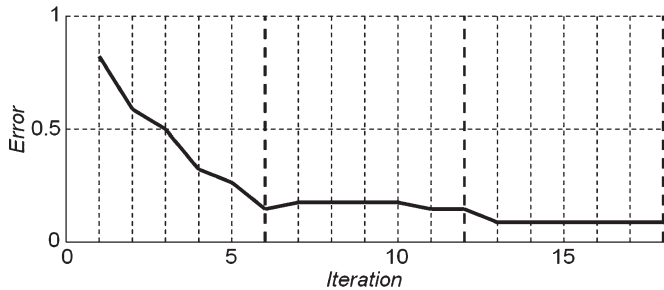


Fig. 16. Training process of F-ECOC that embeds the first-three optimal trees for the speed group.

In the different variants of boosting, we apply 50 iterations. We use Gentle Adaboost, because it has been shown to outperform other Adaboost variants in real applications [44]. With regard to the F-ECOC base classifier, we apply FLDA in the experiments that classify traffic signs and 50 runs of Gentle Adaboost with decision stumps on the UCI data sets. This last choice was selected so that all strategies share the same base classifier.

Table III shows the characteristics of the data used for the classification experiments, where #Training, #Test, #Features, and #Classes correspond to the number of training samples, test samples, features, and classes, respectively.

The classification results and confidence intervals are graphically shown in Fig. 15 for the different groups. It is shown that the F-ECOC that uses FLDA as a base classifier attains the highest accuracy in all cases. Nevertheless, for the circular and triangular signs, the differences among classifiers are significantly different because of the high discriminability of these two groups. The speed group is a more difficult classification problem. In this case, the F-ECOC strategy obtains an accuracy of 90%, outperforming the rest of the classifiers.

3) *Tree-Embedding Analysis*: The training evolution of the F-ECOC at the previous experiment is shown in Fig. 16 for the speed group. Each iteration of the figure shows the classification accuracy by embedding a new node (i.e., a binary classifier) from each optimal tree in the F-ECOC matrix M . The three optimal trees are split by the dark vertical lines. The respective trees are shown in Fig. 17. In the first generated tree in Fig. 17, it is shown that the most difficult partitions are reserved for the final classifiers of the tree. The next trees select the following best partitions of classifiers to avoid repeating classifiers. These classifiers learn subgroups of classes from the same data, improving the classification results (see Fig. 16) by sharing their knowledge among classes.

4) *UCI Evaluation*: To validate the accuracy of the F-ECOC strategy, we tested it on the public UCI Machine Learning Repository [45]. The characteristics of the UCI data sets are shown in Table IV. In this case, to observe the benefits of using multiple-tree embedding, we compared the performance of the F-ECOC strategy with the discriminant ECOC (DECOC) approach [24]. Both methods use 50 runs of Gentle Adaboost with decision stumps as the base classifier, using 2-optimal trees for the F-ECOC strategy.

The classification results of the two strategies over the UCI data sets are shown in Table V. It is observed that the F-ECOC strategy outperforms, in most cases, the results from the DECOC strategy for the same base classifier. Moreover, in the worst case, the F-ECOC methodology obtains the same results as in the DECOC technique.

5) *Model-Fitting Classification*: To test the performance of the classification step of the system, model fitting and F-ECOC classification are applied to a set of 200 regions of interests for each group. The regions of interest are obtained from the detection step. The results are shown in Table VI. It is shown that for circular and speed signs, the results are practically maintained from the previous experiments. For triangular signs, the accuracy is slightly decreased because of the effect of noise, variability of sign appearance, and resolution, which makes the Hough transform lose some sides of the triangular signs. Nevertheless, the final results are more than 90% in all cases.

6) *System Results*: We calculate the performance of the whole system over a test set of 10,000 stereo pairs of images, which correspond to a 100-km road. The accuracy of the real traffic sign recognition system that applies the detection and classification approaches jointly obtains a mean triangular sign reliability of $90.87\% \pm 0.87\%$ and a circular sign reliability of $90.16\% \pm 1.01\%$. In the detection stage, recognition failures are due to background confusion (see Fig. 13) and high interclass variability, whereas in the classification stage, the errors are due to the poor resolution of the images.

7) *Discussions*: Classical approaches to traffic sign detection are based on a segmentation of the image using thresholding over a color space. It is useful, because traffic signs use a color code to easily be identifiable by humans. These colors are used to be distinguished in natural scenes. However, some objects (e.g., vehicles, buildings, and advertising near the road) share similar colors. In addition, the acquired color information is related to the camera that was used, and it is far from the real object's color. Although, in some works, this problem is attenuated using some heuristics (e.g., the size and position of

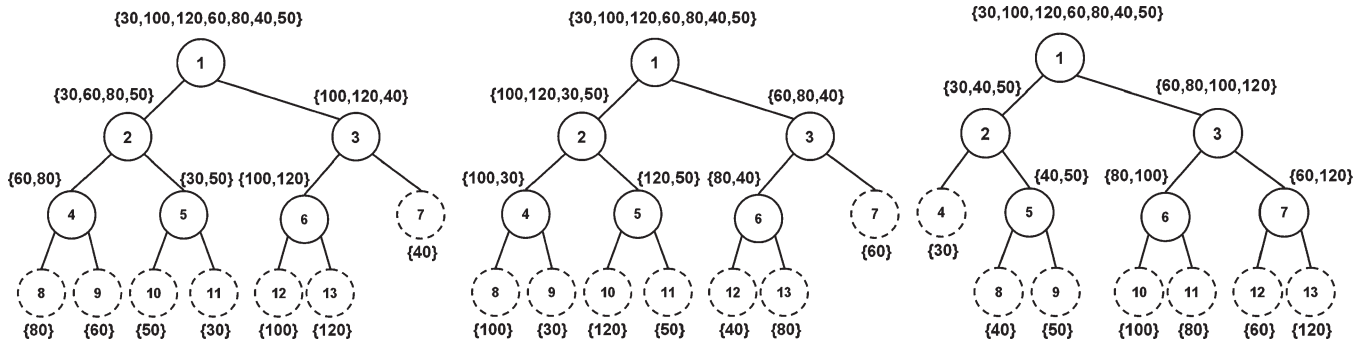


Fig. 17. Three optimal trees generated by F-ECOC for the speed group.

TABLE IV
UCI REPOSITORY DATA SETS CHARACTERISTICS

Problem	#Train	#Attributes	#Classes	Data types	Attribute types	Year
Dermatology	366	34	6	Multivariate	Categorical, Integer	1998
Ecoli	336	8	8	Multivariate	Real	1996
Vowel	990	10	11	Multivariate	Real	-
Yeast	1484	8	10	Multivariate	Real	1996
Satimage	6435	36	7	Multivariate	Integer	1993
Pendigits	10992	16	10	Multivariate	Integer	1998
Segmentation	2310	19	7	Multivariate	Real	1990

TABLE V
CLASSIFICATION RESULTS FROM THE UCI DATA SETS

UCI	Forest ECOC	DECOC
Yeast	53.85±1.64	51.15±2.14
Dermatology	95.32±1.31	93.10±1.81
Ecoli	83.98±1.13	78.15±1.42
Segmentation	94.98±0.66	92.30±0.89
Satimage	73.91±1.11	73.91±1.11
Vowel	77.67±1.81	73.84±1.90
Pendigits	81.42±1.93	81.42±1.93

TABLE VI
MODEL FITTING AND CLASSIFICATION RESULTS

Recognition problem	Accuracy
Circular	98.00
Speed	91.50
Triangular	92.50

the segmented regions or shape), it can be a source of false positives that can interfere on the recognition of the signs.

Our work avoids these problems by using appearance information instead of directly using the gray-scale information. The presented system uses relations between gray values inside the image, and thus, the changes that were produced by different acquisitions systems are smoothed.² However, the appearance-based methods improve the results, but they are not free. The presented methods introduces an extra computation cost to the detection stage, which disable them to be used in driver support applications, where the real-time requirement is harder than in the case of the mobile-mapping problem. To face those restrictions, the detection stage must be optimized. There are

²Note that although color is not required in any stage of the present system, and all the feature sets that were used on the detection stage can be extended to any color space.

promising works that are focused on cascade optimizations [46] and multiple detectors organization [47], which can be used to improve not only the results but the computation time as well. In addition, the original detection scheme of Viola and Jones [21] was successfully embedded in hardware devices, allowing high-detection frame rates. Similar strategies can be adopted for our method.

The experiments that were performed for the F-ECOC classification technique used a greedy search to look for the optimal subgroups of classes that form the tree structures. In our case, the exhaustive search was computationally feasible with the number of classes. If necessary, different strategies can be applied instead. In case of having a large number of classes to be learned, suboptimal solutions can be found using faster approaches (e.g., the Sequential Forward Floating Search [48]) to speed up the method.

In the same way, after a preliminary set of experiments, we fixed FLDA as the base classifier for the F-ECOC strategy to learn the binary problems. It is shown to be a suitable choice for traffic sign classification. In other types of classification problems, the F-ECOC approach can be applied with other base classifiers that better adapt to learn a particular distribution of the data.

The detection and recognition system has been implemented to automatically process hundreds of video sequences that were obtained from the mobile-mapping process in [25]. This data is processed by the Institut Cartogràfic de Catalunya to extract cartographic information. Before the system was implemented, the labeling of the traffic signs in image sequences was manually done by a user. With the automatic system, the recompilation of the information is about three times faster. Furthermore, comparing the results of labeling with the automatic and manual processes, we can argue that the automatic

performance obtains better results. It is done by the fact that the performance of manual labeling is affected by the eyestrain of the user that processes thousands of frames and the extreme illumination conditions of the video sequences.

With regard to the computation cost of the system, the final cascade of detectors learned with the evolutionary Adaboost is comparable in complexity to the Viola and Jones [21] face detector. Moreover, the same optimization technique for calculating the Haar-like features through the integral image can be used in the case of dissociated dipoles. Using ordinal features, the comparison with a float value has been replaced with the use of the sign, which reduces the computation cost compared with the Viola and Jones real-time detector. Regarding the classification stage, although the training time could be expensive, depending on the number of classes, training examples, and the number of features per data sample, once the F-ECOC hypotheses are learned, the classification time depends on the nature of the applied base classifier. In particular, considering FLDA or Adaboost, the classification decision only requires a simple matrix product or an additive model estimation, which can be computed in real time. Nevertheless, although the detection and classification techniques can be optimized to be real-time approaches, the use of correlation on the stereo-association stage, the model fitting, and normalization methodologies prevent the whole system from being set in real time.

With regard to the memory usage, the final detection system needs eight integers per feature to represent the dissociated dipoles. During detection, at least the integral image of the processed image must be in memory. Finally, the classification parameters and the ECOC matrix must also be available in memory. In summary, the stronger restriction in memory usage is the size of processed images and not the system itself.

V. CONCLUSION

We have presented a mobile mapping detection and classification system to deal with the traffic-sign-recognition problem. We have introduced a computationally feasible method for feature selection based on an evolutionary strategy. The exhaustive search over all the combinations between feature and threshold are replaced with the evolutionary approach and ordinal features; thus, the final approach has two main advantages: 1) It speeds up the learning process, and 2) it allows for working with large feature sets to distinguish between object and background, which is computationally unfeasible using traditional methods because of the large number of features. Moreover, we have proposed the F-ECOC classification strategy to lead with the multiclass categorization problem. The method is based on the embedding of multiple optimal tree structures in the ECOC framework, sharing their knowledge among classes, and constructing an ensemble of trees until the necessary performance is achieved. A wide set of traffic signs are recognized under high variance of appearance, such as noise, affine deformation, partial occlusions, and reduced visibility. The validation of the mobile mapping system shows that the presented methodology offers high robustness and better performance compared with the state-of-the-art strategies in uncontrolled environments.

REFERENCES

- [1] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. von Seelen, "An image processing system for driver assistance," in *Proc. IEEE Int. Conf. IV*, Stuttgart, Germany, 1998, pp. 481–486.
- [2] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Paetzold, and C. Wohler, "Autonomous driving goes downtown," *IEEE Intell. Syst.*, vol. 13, no. 6, pp. 40–48, Nov/Dec. 1998.
- [3] S. Madeira, L. Bastos, A. Sousa, J. Sobral, and L. Santos, "Automatic traffic signs inventory using a mobile mapping system," in *Proc. Int. Conf. Exhib. Geogr. Inform. GIS PLANET*, Lisboa, Portugal, 2005.
- [4] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proc. IEEE Intell. Veh. Symp.*, 2004, pp. 566–571.
- [5] H. Akatsuka and S. Imai, "Road signposts recognition system," in *Proc. SAE Veh. Highway Infrastructure—Safety Compatibility*, 1987, pp. 189–196.
- [6] D. Kellmeyer and H. Zwahlen, "Detection of highway warning signs in natural video images using color image processing and neural networks," in *Proc. IEEE Int. Conf. Neural Netw.*, 1994, vol. 7, pp. 4226–4231.
- [7] M. de Saint Blancard, "Road sign recognition: A study of vision-based decision making for road environment recognition," in *Vision-Based Vehicle Guidance*. New York: Springer-Verlag, 1992, pp. 162–172.
- [8] A. D. L. 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.
- [9] D. Ghica, S. W. Lu, and X. Yuan, "Recognition of traffic signs using a multilayer neural network," in *Proc. Can. Conf. Elect. Comput. Eng.*, Halifax, NS, Canada, 1994, vol. 44, pp. 848–859.
- [10] A. de la Escalera, J. M. Armingol, and M. Salichs, "Recognition of traffic signs using a multilayer neural network," in *Proc. 3rd Int. Conf. Field Service Robot.*, Espoo, Finland, Jun. 2001, pp. 833–834.
- [11] J. Miura, T. Kanda, and Y. Shirail, "An active vision system for real-time traffic sign recognition," in *Proc. IEEE Int. Conf. ITSC*, 2000, pp. 52–57.
- [12] D. Shaposhnikov, L. Podladchikova, A. Golovan, N. Shevtsova, K. Hong, and X. Gao, *Road Sign Recognition by Single Positioning of Space-Variant Sensor*, 2002. [Online]. Available: citeseer.csail.mit.edu/657126.html
- [13] S. Hsu and C. Huang, "Road sign detection and recognition using matching pursuit method," *Image Vis. Comput.*, vol. 19, no. 3, pp. 119–129, Feb. 2001.
- [14] G. Piccioli, E. D. Micheli, and M. Campani, "A robust method for road sign detection and recognition," in *Proc. ECCV*, 1996, vol. 1, pp. 495–500. [Online]. Available: citeseer.csail.mit.edu/piccioli96robust.html
- [15] G. Piccioli, E. D. Michelli, P. Parodi, and M. Campani, "Robust road sign detection and recognition from image sequences," in *Proc. Intell. Veh.*, 1994, pp. 278–283.
- [16] S. Escalera and P. Radeva, "Fast gray scale road sign model matching and recognition," in *Recent Advances in Artificial Intelligence Research and Development*. Amsterdam, The Netherlands: IOS, Oct. 2004.
- [17] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.
- [18] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 264–278, Jun. 2007.
- [19] A. de la Escalera, J. Armingol, J. Pastor, and F. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 2, pp. 57–68, Jun. 2004.
- [20] P. Paclik, J. Novovicova, and R. Duin, "Building road-sign classifiers using a trainable similarity measure," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 309–321, Sep. 2006.
- [21] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [22] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. IEEE Conf. Image Process.*, 2002, pp. 155–162.
- [23] T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *J. Artif. Intell. Res.*, vol. 2, pp. 263–286, 2005.
- [24] O. Pujol, P. Radeva, and J. Vitrià, "Discriminant ECOC: A heuristic method for application dependent design of error-correcting output codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 1001–1007, Jun. 2006.
- [25] R. Alamús, A. Baron, E. Bosch, J. Casacuberta, J. Miranda, M. Pla, S. Sanchez, A. Serra, and J. Talaya, "On the accuracy and performance of the geomobile system," in *Proc. ISPRS*, Istanbul, Turkey, Jul. 2004.

- [26] K. Thoresz and P. Sinha, "Qualitative representations for recognition," *J. Vis.*, vol. 1, no. 3, p. 298, Dec. 2001. [Online]. Available: <http://journalofvision.org/1/3/298/>
- [27] B. J. Balas and P. Sinha, "STICKS: Image-representation via non-local comparisons," *J. Vis.*, vol. 3, no. 9, p. 12, Oct. 2003.
- [28] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded-up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, May 2006, pp. 404–417.
- [29] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Mach. Learn.*, 1996, pp. 148–156.
- [30] Y. Freund and R. Schapire, "A short introduction to boosting," *J. Jpn. Soc. Artif. Intell.*, vol. 14, no. 5, pp. 771–780, Oct. 1999.
- [31] T. Dietterich and E. Kong, "Error-correcting output coding corrects bias and variance," in *Proc. 12th Int. Conf. Mach. Learn.*, S. Prieditis and S. Russell, Eds., 1995, pp. 313–321.
- [32] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, 2002.
- [33] N. J. Nilsson, *Learning Machines*. New York: McGraw-Hill, 1965.
- [34] T. Hastie and R. Tibshirani, "Classification by pairwise grouping," in *Proc. NIPS*, 1998, vol. 26, pp. 451–471.
- [35] S. Escalera, O. Pujol, and P. Radeva, "Decoding of ternary error-correcting output codes," in *Proc. CIARP*, 2006, pp. 753–763.
- [36] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.
- [37] B. S. Morse, "Segmentation (edge based, Hough transform)," Brigham Young Univ., Provo, UT, 2000. Tech. Rep.
- [38] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner-Verlag, 1998.
- [39] P. Simard, Y. LeCun, J. Denker, and B. Victorri, "Transformation invariance in pattern recognition: Tangent distance and tangent propagation," in *Proc. Neural Netw.—Tricks of the Trade*, 1998, vol. 1524, pp. 239–274.
- [40] S. Dudoit, J. Fridlyand, and T. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," Dept. Stat., Univ. California Berkeley, Berkeley, CA, Tech. Rep. 576, Jun. 2000.
- [41] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, 2002. Tech. Rep.
- [42] Y. Kim, S. Hahn, and B. Zhang, "Text filtering by boosting naive Bayes classifiers," in *Proc. SIGIR Conf. Res. Develop.*, 2000, pp. 168–175.
- [43] A. Torralba and K. Murphy, "Sharing visual features for multiclass and multiview object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 854–869, May 2007.
- [44] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," Dept. Stat., Stanford Univ., Palo Alto, CA, 1998. Tech. Rep.
- [45] A. Asuncion and D. Newman, *UCI Machine Learning Repository*, 2007.
- [46] H. Luo, "Optimization design of cascaded classifiers," in *Proc. Comput. Vis. Pattern Recog.*, 2005, vol. 1, pp. 480–485.
- [47] C. Huang, H. Ai, Y. Li, and S. Lao, "Vector boosting for rotation invariant multiview face detection," in *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 446–453.
- [48] P. Pudil, F. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. Int. Conf. Pattern Recog.*, 1994, pp. 279–283.



Sergio Escalera received the B.S., M.S., and Ph.D. degrees from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2003, 2005, and 2008, respectively.

He is a Collaborating Professor with the Department of Applied Mathematics and Analysis, Universitat de Barcelona. His research interests include machine learning, statistical pattern recognition, and visual object recognition.



Jordi Vitrià received the Ph.D. degree from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 1990. His Ph.D. dissertation was focused on mathematical morphology and image analysis.

In 1986, he joined the Computer Science Department, UAB, where he has been an Associate Professor since 1991. In 2007, he joined the Department of Applied Mathematics and Analysis, Universitat de Barcelona. He has taught courses on various topics in artificial intelligence and computer vision, including undergraduate and graduate courses.

Since 1991, he has organized several activities on image analysis, computer vision, and artificial intelligence. He has published a number of papers on pattern recognition and computer vision. His research interests include the application of machine learning techniques to visual object recognition problems and the development of statistical machine learning techniques.



Oriol Pujol received the Ph.D. degree in computer science from the Universitat Autònoma de Barcelona, Barcelona, Spain, in 2004.

He is currently a Lecturer with the Department of Applied Mathematics and Analysis, Universitat de Barcelona. His research interest includes basic statistical machine learning techniques for object recognition and medical imaging analysis.



Xavier Baró received the B.S. and M.S. degrees in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2003 and 2005, respectively. He is currently pursuing the Ph.D. degree, under the supervision of Dr. J. Vitrià, with the Computer Vision Center, UAB.

His research interests include machine learning, evolutionary computation, and statistical pattern recognition.



Petia Radeva received the Ph.D. degree from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. Her Ph.D. dissertation was focused on the development of physics-based models applied to image analysis.

She is currently an Associate Professor with the Department of Computer Science, UAB. In 2007, she has joined the Department of Applied Mathematics and Analysis, Universitat de Barcelona. Her research interests include the development of physics-based and statistical approaches for object recognition,

medical image analysis, and industrial vision.