Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 1759–1768
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Boosted Landmarks of Contextual Descriptors and Forest-ECOC: A novel framework to detect and classify objects in cluttered scenes

Sergio Escalera [a,*], Oriol Pujol [b], Petia Radeva [a]

[a] Centre de Visió per Computador, Campus UAB, 08193 Bellaterra, Barcelona, Spain
[b] Dept. Matemática Aplicada i Anàlisi, UB, Gran Via 585, 08007 Barcelona, Spain

Received 25 April 2006; received in revised form 28 February 2007
Available online 25 May 2007

Communicated by F. Roli

Abstract

In this paper, we present a novel methodology to detect and recognize objects in cluttered scenes by proposing boosted contextual descriptors of landmarks in a framework of multi-class object recognition. To detect a sample of the object class, Boosted Landmarks identify landmark candidates in the image and define a constellation of contextual descriptors able to capture the spatial relationship among them. To classify the object, we consider the problem of multi-class classification with a battery of classifiers trained to share their knowledge among classes. For this purpose, we extend the Error Correcting Output Codes technique proposing a methodology based on embedding a forest of optimal tree structures. We validated our approach using public data-sets from the UCI and Caltech databases. Furthermore, we show results of the technique applied to a real computer vision problem: detection and categorization of traffic signs.

Keywords: Object detection; Object recognition; Boosting; Correlogram; Multiple classifiers; Embedding of dichotomies

1. Introduction

Usually, the problem of object recognition (e.g. person identification) needs a previous detection of the object category (e.g. face location). Object detection is concerned with the reliable and accurate location of target objects in an image. According to the way objects are described, three main families of approaches can be considered (Murphy et al., 2003): part-based, patch-based and region-based methods. Part-based approaches consider that an object is defined as a specific spatial arrangement of its parts fragments. Following this idea, an efficient Bayessian network for learning the spatial arrangement of parts is proposed in (Schneiderman, 2004). A unsupervised statistical learning of constellation of parts and spatial relations is used in (Fergus and Zisserman, 2003). Other authors (Hong and Huang, 2004) propose to use Attribute Relational Graphs for describing spatial relations. In (Amores et al., 2005) a representation integrating Boosting with constellations of contextual descriptors is defined. In this work, the feature vector includes the bins that correspond to the different positions of the correlograms determining the object properties. Another family of recognition techniques is the patch-based methods, which classify each rectangular image region of a fixed aspect ratio (shape) at multiple sizes, as object (or parts of the target object) or background. In this topic, the authors of Agarwal et al. (2004) use a dictionary of parts and a window algorithm for learning active features of the object are proposed. A similar technique is found in (Torralba et al., 2004), where objects are described by the best features obtained using

* Corresponding author. Tel.: +34 68 729 19 57; fax: +34 93 581 16 70.
E-mail addresses: sescalera@cvc.uab.es (S. Escalera), oriol@cvc.uab.es (O. Pujol), petia@cvc.uab.es (P. Radeva).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.05.007

masks and normalized cross-correlation. Finally, region-based algorithms segment regions of the image from the background and describe them by a set of features that provide texture and shape information. The selection of feature points can be based on image contour points (Amores et al., 2005) or other image features.

Once an object is located in an image, it should be recognized using some kind of classification technique (support vector machines, nearest neighbor, linear discriminant analysis, etc.). Recently, Torralba et al. (2004) proposed a novel multi-class approach where instead of training independent classifiers for each object class, they are jointly trained leading to a more robust feature extraction and better recognition generalization. Following the multi-class framework, we choose to use the Error Correcting Output Codes (ECOC) (Dietterich and Bakiri, 2005) technique. This technique shown to be a very successful multi-class framework due to its ability to extend any binary classifier to the multi-class classification domain. However, the ECOC design is still an open issue. Recently, embedding of a tree structure in the ECOC framework has been shown to obtain high accuracy with a very small number of binary classifiers (Pujol et al., 2006). Here, we take advantage of the representation of tree structures in the ECOC framework to introduce a "Forest"-ECOC. This novel method is based on embedding of different optimal trees in the ECOC approach to obtain the necessary number of classifiers assuring the required classification performance.

Our goal in this paper is 2-fold: first, we introduce a novel approach for detection of objects in cluttered scenes based on Boosted Landmarks to identify landmark candidates in the image. The features used are invariant to global illumination, to slight image transformation and partial occlusions. On the other hand, according to the landmark candidates, a constellation of contextual descriptors using correlograms is defined for each landmark to capture the spatial relationship between them. Using boosting to learn the object descriptors ensures a good theoretical and practical convergence to a low recognition error rate in few iterations. Second, a new multi-class learning technique is introduced based on embedding a forest of optimal trees in an ECOC framework that allows to share classifiers (tree nodes, base classifiers or dichotomies) across classes in a very robust way. The main advantage of the proposed classification technique is the problem-dependent design, leading to a compact codeword with high generalization performance power. Different experiments are evaluated on synthetic and real data, showing the high performance of the Boosted Landmarks of Contextual Descriptors and Forest-ECOC approaches.

The article is organized as follows: Section 2 provides a description of the proposed object detection method. Section 3 introduces the novel Forest-ECOC approach. Section 4 shows the results of our method on different databases and on a real problem of traffic sign detection and recognition. Finally, Section 5 concludes the paper.

2. Object detection by Boosted Landmarks and contexts

In this section, we introduce a new object detection method based on training the discriminant features of the object description. Such description includes the information of correlograms to learn at the same time the object local representation and the spatial relationship among its parts fragments.

2.1. Patch-based step: Boosting Landmarks

A common strategy to address the object detection problem is to model the object as an arrangement of its parts. The representative parts of the object (e.g. represented by a set of landmarks) must be highly discriminable; incorporating the spatial relationship between different parts (Belongie et al., 2000) can improve significantly the robustness of the object detection.

In order to avoid considering all possible ROIs of an image where an object can be located, first we define candidate locations of the object of interest by means of a set of landmarks. The set of object landmarks is selected manually from a data set. Using a training set of positive samples and a negative set of background image regions, we train each landmark using a cascade of classifiers (Hastie and Tibshirani, 1998). In particular, Gentle Adaboost with Haar-like features estimated on the Integral Image (Baro and Vitria, 2004) has been used in the cascade since it has been shown to outperform most of the other boosting variants in real applications (Friedman et al., 1998). Each level of the cascade is specialized on a complex set of features corresponding to a landmark. By adding cascade levels, the number of false positives is reduced while maintaining the detection of true positives, and the process is repeated for each landmark of the object. This approach has the advantage of reducing the number of landmark candidates when compared to other well-known techniques. For instance, Torralba et al. (2004) use a set of masks and parts of an object and use normalized cross-correlation to obtain and detect the set of landmarks. By using the Haar-like features, compared to other methods like the normalized cross-correlation (Torralba et al., 2004), we are more permissive to detect objects in case of object transformations and to obtain a lower level of confusion with the background regions. Summarizing, the steps to train a landmark detector are

For each landmark:

- Define a positive set of image regions (centered in the landmark).
- Define a set of non-containing landmark images (negative set).
- Train a cascade of classifiers for each landmark.

To illustrate the process observe the triangular traffic sign image in Fig. 2a. To distinguish this object type, we have manually identified six different object parts (land-

marks) that can represent the object. The selected fragments are shown in Fig. 1. In the detection step, the set of selected landmarks is learnt using Gentle Adaboost with the Haar-like features estimated in the integral image. In particular, for the example in Fig. 2 we used 100 real triangular signs to generate a set of 100 positive samples of 21×21 pixels for each landmark. For each fragment, its 100 positives samples and 500 random background samples of the same size are trained in an attentional cascade of 10 levels, allowing a false alarm rate by stage of 30%. This measure assures that each landmark classifier has learnt correctly 100% of the positives samples, and the small number of detected false positives does not introduce ambiguity at the detection step. The use of six landmark cascades gives the results shown in Fig. 2b. We can observe that it has a small number of detected labelled landmarks compared to all possible locations and scales. Note that the presented scheme is quite robust to scale, translation, global illumination and to small object affine transformations, avoiding the problems of background confusion of masked landmarks because of the use of the Haar-like features (Baro and Vitria, 2004).

2.2. Parts-based step: Contextual descriptors

In order to refine the set of landmarks we use their contextual description. This step focuses on defining the spatial relationship among the previously detected landmarks to be learnt. Our approach proposes an alternative point of view of the method of Amores et al. (2005) in which a set of points of interest $P = \{p_i\}_{i=1}^{N}$ is considered, where N is the number of points of interest coming from the edges of the image. These points are used to build the constellation of multi-scale correlograms. However, opposed to the work presented in (Amores et al., 2005), our relevant information is provided by landmarks instead of a set of contour points. Since we are focusing on landmark candidates, we can exploit the previous knowledge about the relationship between the landmarks and the size of the object of interest, reducing considerably the number of false positives and avoiding the multi-scale step. Considering n landmarks and their sets of detected candidates $L^1 = \{L_1^1 \dots L_{i_1}^1\}, \dots, L^n = \{L_1^n \dots L_{i_n}^n\}$, where i_j is the number of instances of landmark j found in the image, for each combination of possible landmarks candidates $\{L_{j_1}^1, \dots, L_{j_n}^n, j_1 \in \{1, \dots, i_1\}, \dots, j_n \in \{1, \dots, i_n\}\}$, we generate n correlograms centered at the n chosen candidates. Their combination forms a constellation. From this constellation, we design a contextual descriptor vector $D = (D_1, \dots, D_n)$. The descriptor vector associated to each landmark candidate is described by $D_i = \{B_i^1, \dots, B_i^n\}$, being $B_i^j = \{(o_j, h_j, x_j)\}_{i=1}^{n}$, where o_j is the label identifying the part, h_j are the properties describing the part, and x_j is its spatial description in the image defined by its shape context (Belongie et al., 2002), as shown in Fig. 2c. Hence, the spatial relationship vector is defined by the values of the correlogram bins for each of the landmarks. For example, using the six landmarks shown in Fig. 1, the spatial descriptor vector for an object is $6 \times N$ bins in length, where N is the number of bins that forms each correlogram.

Given L correlograms of N bins, we create the object descriptor rearranging the bins as a vector of size $N \times L$. Since the constructed descriptor is, usually, very highly dimensional, we use Gentle Adaboost (as in the case of learning the landmarks) as a feature selection algorithm to reduce the dimensionality of the feature space and to learn the representative features of the object. In this way, the final classifier learns at the same time the features that correspond to relevant landmarks and their respective spatial relations. Note that we can also introduce extra information, such as the image contour map as an additional information to the Boosted Landmarks.

The detected landmarks involved in the detection step are shown in Fig. 2b. First, all candidates of each type of

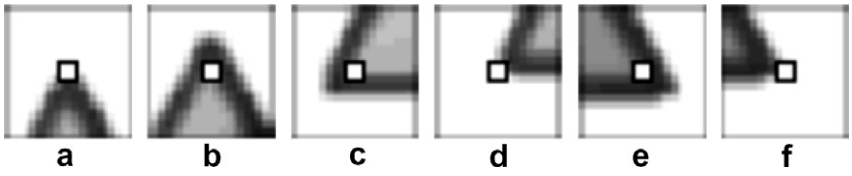


Fig. 1. Selected landmarks for triangular signs.

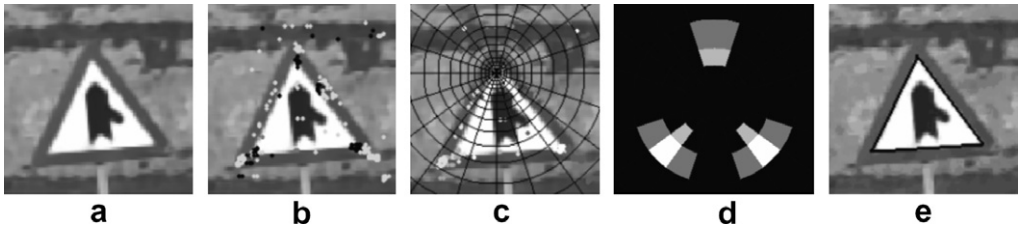


Fig. 2. (a) Input image. (b) Detected landmarks. (c) Contextual descriptors. (d) Resulting bins at feature selection of the correlogram of the landmark of Fig. 1b. (e) Detected sign.

landmark (each positive detection of a given landmark) are sorted by their likelihood using the margin of the output of the Gentle Adaboost classifier. Afterwards, we select the first combination of landmarks (one of each type) that maximizes the sum of the likelihood. The individual vector descriptors of each set of selected landmarks are merged. In Fig. 2c, the correlogram applied to the landmark displayed in Fig. 1b is shown. In Fig. 2d, the locations learnt for the correlogram of the same landmark are shown. The gray level of the bins of the correlogram corresponds to the importance assigned by the boosting procedure – which is intuitively related to the likelihood of the presence of the other landmarks in the descriptor of a current landmark. When a combined descriptor from a set of landmark candidates is classified as positive using the trained Gentle Adaboost classifier, the object presence and the location of its landmarks are defined. In Fig. 2e we can observe a detected object, which contextual descriptor, defined by the combination of detected landmarks, has been accepted as a positive example using the classifier based on Boosted Landmarks.

3. Object recognition by Forest-ECOC

Once located the object category (e.g. a traffic sign), we proceed with the multi-class object recognition. The recognition is handled using a multi-class framework based on ECOC. The basis of this framework is to create a compact codeword for each of N_c classes (generating N_c codewords, respectively). Arranging the codewords as rows of a matrix, a "coding matrix" M is defined where $M_{N_c \times n} = \{m\}$, $m \in \{-1, 0, 1\}$, where n is the length of the code representing each class. Each column of the coding matrix is determined by a binary classifier (dichotomy). From the point of view of learning, the coding matrix M can be seen as the process of embedding n binary learning problems corresponding to the n columns of the matrix-coding each class with $+1$, 0 and -1 according to their class membership. A zero value indicates that a particular class is not considered for a given binary classifier. Given a new data description, a code is obtained for each data point in the test set as a result of the outputs of the n dichotomies. This is compared with the N_c codewords – corresponding to the matrix rows – and it is assigned to the class with the "closest" codeword.

In order to design an ECOC system, we need a coding and a decoding strategy. When the ECOC technique was first developed (Dietterich and Bakiri, 1991) it was believed that the ECOC code matrices should be designed to have certain properties to enable them to generalize well. A good error-correcting output code for a k-class problem should satisfy that rows and columns, and their complementaries are well-separated from the rest in terms of Hamming distance. Most of the discrete coding strategies up to now are pre-designed problem-independent codewords satisfying the requirements of row and column separability. However, our Forest-ECOC technique design is problem-dependent generating as much dichotomies as necessary to obtain the required performance independently of the kind of classifiers used.

Concerning the decoding strategies, two of the most standard techniques are the Euclidean distance $d_j(x, y^j) = \sqrt{\sum_{i=1}^{n}(x_i - y_i^j)^2}$ and the Hamming decoding distance $d_j(x, y^j) = \sum_{i=1}^{n}|(x_i - y_i^j)|/2$, where d_j is the distance to the row j, n is the number of dichotomies, and x and y^j are the values of the test input vector codes and the jth base class codeword, respectively. The benefit of ECOCs consists in the fact that if the minimum Hamming distance between any pair of codewords is d, then any $[(d-1)/2]$ errors in the individual dichotomies classification can be corrected, keeping as the nearest codeword the correct codeword.

The analysis of the ECOC error evolution has demonstrated that the ECOC corrects errors caused by the bias and the variance of the learning algorithm (Dietterich and Kong, 1995).[1] The variance reduction is to be expected, since ensemble techniques address this problem successfully and ECOC is a form of voting procedure. On the other hand, the bias reduction must be interpreted as a property of the decoding step. It follows that if a point x is misclassified by some of the learnt dichotomies, it can still be classified correctly after being decoded due to the correction ability of the ECOC algorithm. Non-local interaction between training examples leads to different bias errors. Initially, the experiments in (Dietterich and Kong, 1995) show the bias and variance error reduction for algorithms with *global* behavior (when the errors made at the output bits are not correlated). After that, new analysis also shows that ECOC can improve performance of *local* classifiers (e.g. the k-nearest neighbor, which yields correlated predictions across the output bits) by extending the original algorithm or selecting different features for each bit (Ricci and Aha, 1998).

In (Pujol et al., 2006), a method for embedding a tree structure in the ECOC framework is proposed. Beginning from a root of a tree – that considers all the classes of the problem – a binary tree is built as follows. Each node corresponds to the best bi-partition of the set of classes maximizing the quadratic mutual information between the class samples and their labels. The process is recursively applied until sets of single classes corresponding to the tree leaves are obtained. Taking this work as a baseline, we propose the embedding of multiple trees to form a Forest-ECOC. However, opposed to the discriminant tree proposed in (Pujol et al., 2006), we use the classification score to create each node of the tree. The tree with the maximum classification score at each node is called "optimal" tree. In the case that we consider the first T best partitions of clas-

[1] The bias term describes the component of the error that results from systematic errors of the learning algorithm. The variance term describes the component of the error that results from random variation and noise in the training samples and random behavior of the learning algorithm. For more details, see (Dietterich and Kong, 1995).

Table 1
Training algorithm for the Forest-ECOC

Given n classes: $C_1,\ldots,C_n$ and T the number of optimal tree structures to be embedded:

Step 1. Initialize the root node with the set $N_0 = \{C_1,\ldots C_n\}$

Step 2. Generate the tree structures:

- For each node N_j consider the T partitions $\wp_{kj} = \{\{\wp_j^1, \wp_j^2\} | N_j = \wp_j^1 \bigcup \wp_j^2\}, k = 1 \ldots T$ that attain the minimal empirical error for the subproblem defined by the partition $\wp_{kj}$

$$\wp_k = \underset{\tilde{\wp}}{\operatorname{argmin}}(e(\mathscr{H}(\tilde{\wp}, \mathbf{x}), \mathbf{l})) \tag{1}$$

  where $e(H(\cdot, x), l)$ stands for the empirical error between the hypothesis result $H(\cdot, x)$ on the data set x and the respective class labels l
- Partitions $\wp_{kj}$, $k = 2,\ldots,T$ define $T-1$ roots of new trees of the forest
- Include each binary classifier h_j for each internal node of the trees as a column in the Forest-ECOC matrix M, using the following rule for each class C_r:

$$M(r,j) = \begin{cases} 0 & \text{if } C_r \notin N_j \\ +1 & \text{if } C_r \in \wp_j^1 \\ -1 & \text{if } C_r \in \wp_j^2 \end{cases} \tag{2}$$

ses – the T partitions that best splits the set of classes – it allows us to create multiple tree structures. These trees are embedded in the ECOC matrix forming the Forest-ECOC. The pseudo-algorithm is shown in Table 1.

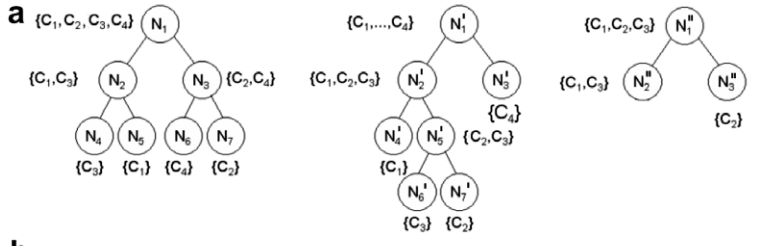
For a given multi-class problem, we have seen experimentally that usually using 2 or 3 trees is enough to create a rich set of dichotomizers to achieve accurate results. An example of the forest for $T = 2$ and the Forest-ECOC matrix for a toy problem are shown in Fig. 3. The two first optimal tree structures (without repeating classifiers) are shown in Fig. 3 for a four-class toy problem. The best 2 bi-partitions for the root are $\{\{C_1, C_3\}, \{C_2, C_4\}\}$ and $\{\{C_1, C_2, C_3\}, \{C_4\}\}$ that correspond to classifiers h_1 and h_4, respectively. For the next nodes of the first optimal tree, we can only generate one bi-partition for each case, corresponding to h_2 and h_3, respectively. For the node N_2' of the second three, the $T = 2$ first best bi-partitions are $h_5 = \{\{C_1\}, \{C_2, C_3\}\}$ and $h_7 = \{\{C_1, C_3\}, \{C_2\}\}$. Note that the classes are shared by different partitions of sets and they are combined in the Forest-ECOC matrix.

Given an input sample to test with the Forest-ECOC matrix, we obtain the Forest-ECOC vector where each vector component is the result of solving each binary classifier trained on each of the columns of the matrix. Note that this procedure can be cast in the multi-task framework since it combines knowledge from different binary problems and shares their knowledge among the tasks.

The second step of the ECOC process is the decoding. When we decode a new test data point, the ambiguity of the zero values can produce an accumulation of errors in the distance estimation – recall that symbol zero is used for all classes not considered in a dichotomy. The main drawback of the traditional decoding strategies is the ambiguity that appears due to the zero values (Pujol et al., 2006). In order to address this issue, we base our classification on the Euclidean distance $d_j = \sqrt{\sum_{i=1}^{n}(x_i - y_i^j)^2}$ – that shows to perform the Hamming results when applied to ternary ECOC symbol-based (Pujol et al., 2006), where d_j is the distance to the row j, n is the number of dichotomies, and $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1^j, x_2^j, \ldots, x_n^j)$ are the results of classification of a test example and base codeword of class j, respectively.

4. Results

Given both parts of our approach – object detection and multi-class categorization – in order to show the behavior



b

	h_1	h_2	h_3	h_4	h_5	h_6	h_7
C_1	1	1	0	1	1	0	1
C_2	-1	0	1	1	-1	-1	-1
C_3	1	-1	0	1	-1	1	1
C_4	-1	0	-1	-1	0	0	0

Fig. 3. (a) $T = 2$ bi-partitions of trees for a toy problem, (b) the Forest-ECOC matrix, where h_1, h_2 and h_3 correspond to classifiers of N_1, N_2 and N_3 from the first tree, h_4, h_5 and h_6 to N_1', N_2' and N_5' from the second tree, and h_7 to N_1'' from the third tree.

and to validate the proposed methodology, we designed three types of experiments. First, we compare the Boosted Landmarks of Contextual Descriptors with two state-of-the-art detection approaches on a set of objects images from the public Caltech repository databases. Second, we validate the recognition approach based on the Forest-ECOC technique compared to well-known state-of-the-art multi-classification strategies on a set of public UCI repository databases. Finally, we apply the whole detection and classification system to a real multi-class traffic sign detection and categorization problem.

4.1. Evaluating Boosted Landmarks in Contextual Descriptors

In order to compare the accuracy of our detector, we test the Boosted Landmarks of Contextual Descriptors approach on the Caltech database (www.vision.caltech.edu/html-files/archive.html) considering the following seven object categories: car side, face, motorbike, car rear, plane, leaf, and spotted cat (Fig. 4), training only three landmarks from the models of each database. In Figs. 5 and 6, the models, contour points, landmarks trained, and a correlogram for the face and car side databases are shown. To validate the method we used 20% of samples to train landmarks (between 30 and 80 samples for each category) and contextual descriptors by boosting, and the rest to test. From the 20% images for training, we select only three representative landmarks in a supervised way to train each database (Fig. 4(down)). We use 40 weaks of Gentle Adaboost with Decision Stumps to train the cascades and the correlogram descriptors.

The results of this experiment are shown in Table 2. We compare the results with those reported by Fergus and Zisserman (2003) and the boosting context proposed by Amores et al. (2005). We can see that our proposed technique obtains better results in most of the cases: car (side), face, car (rear), and leaf, and comparable results in the other three cases.

Boosting Context (Amores et al., 2005) shows very good behavior too, but it is more susceptible to confusion and appearance of false positives and negatives due to the use

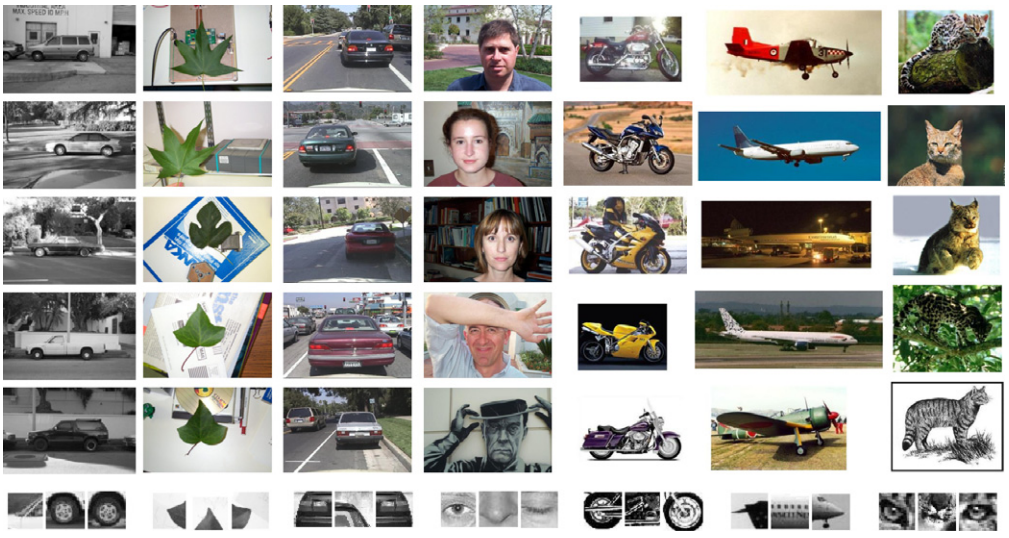


Fig. 4. Some samples for the considered Caltech categories and relevant landmarks trained.

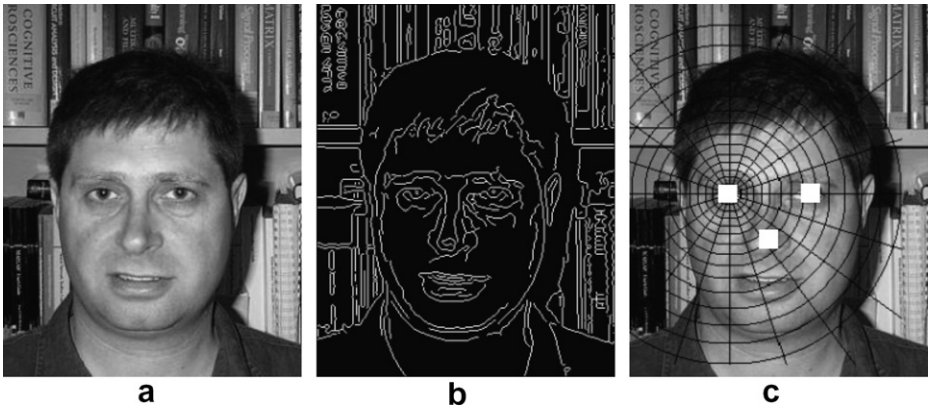


Fig. 5. Fergus faces database. (a) Original image. (b) Contour points map. (c) Correlogram for a given landmark.

Fig. 6. Fergus car side database. (a) Original image. (b) Contour points map. (c) Correlogram of a landmark.

Table 2
Hit ratio results for the Fergus database

Category	Fergus (Fergus and Zisserman, 2003)	Boosting Context (Amores et al., 2005)	Boosted Landmarks in Contextual Descriptors
Car (side)	88.50%	90.00%	**96.63%**
Face	96.40%	89.50%	**97.72%**
Motorbike	92.50%	**95.00%**	93.85%
Car (rear)	90.30%	96.90%	**99.35%**
Plane	90.20%	**94.50%**	92.50%
Leaf	–	96.30%	**98.85%**
Spotted car	**90.00%**	86.50%	84.00%
Rank	2.50	1.86	**1.57**

of the contour points. The authors in (Fergus and Zisserman, 2003) use a model that involves a considerable number of features, being susceptible to false positives appearance. We also tested the false alarm rate using the background set of images from the Caltech database. Testing with 500 background images, our Boosted Landmarks classifiers obtained a maximum on only one false positive at each of the seven object categories.

4.2. Evaluating Forest-ECOC

In order to validate the accuracy of the Forest-ECOC we tested it on the UCI repository. The compared methods are 40 runs of multi-class joint boosting with decision stumps (Torralba et al., 2004),[2] all pairs Fisher Linear Discriminant Analysis (FLDA) with a previous 99.9% of the Principal Components Analysis, and Dense Random ECOC. Our method and Dense Random ECOC use Gentle Adaboost with decision stumps as a classification technique to estimate the Forest-ECOC dichotomies, with $T = 2$ to generate and embed multiple trees. The state-of-the-art in random strategies are the dense random and the sparse random coding techniques. As the dense random strategy tends to improve the classification rate of sparse case for the same number of binary problems (Pujol et al., 2006), we tested this strategy with the same number of dichotomies as our Forest-ECOC approach. The probability of appearance of the $\{1, -1\}$ is 0.5 in both cases, so we tested 10000 matrices to obtain the one that maximizes the row and column Hamming distance (Allwein et al., 2002).

Looking at the results in Table 3 we can observe that our method is competitive with the three commented approaches, and it attains the first position in the classification ranking for 8 UCI databases. The table shows the mean accuracy using stratified 10-fold cross-validation, and the confidence interval at 95% using a two tailed t-test. The ranking has been obtained considering that all techniques with results overlapping with the confidence interval of the top performance technique are considered also as first choice. Observe that the Forest-ECOC compares favorably to the other approaches; in this sense, it turns out a promising technique for the purposes of multi-class recognition.

4.3. Evaluating the whole system in a real traffic sign problem

In order to validate the Boosted Landmarks and the Forest-ECOC detection and classification approaches in a complex real problem, we tested it in a real traffic sign recognition problem.

We used a database of 300 traffic sign images obtained by a mobile mapping process in non-controlled outdoor conditions (Escalera and Radeva, 2004). The considered traffic sign classes are shown in Fig. 7. Some samples of our database illustrating the variation of appearance of the signs are shown in Fig. 8. Observe the high variability of the signs due to the non-controlled conditions of acquisition. Triangular signs are detected using the Boosted

Table 3
Classification results for UCI databases

UCI	JB	All pairs FLDA	Forest ECOC	Dense ramdom ECOC
Yeast	**56.54 ± 1.42**	52.32 ± 1.65	**53.85 ± 1.64**	47.32 ± 0.93
Dermathology	**96.14 ± 0.92**	**96.40 ± 1.33**	**95.32 ± 1.31**	**96.57 ± 0.74**
Ecoli	**85.50 ± 1.06**	**84.62 ± 1.92**	**83.98 ± 1.13**	81.15 ± 1.55
Segmentation	92.83 ± 1.01	86.81 ± 0.91	**94.98 ± 0.66**	73.89 ± 0.56
Satimage	**80.02 ± 1.18**	**81.92 ± 1.92**	73.91 ± 1.11	72.85 ± 0.83
Vowel	64.86 ± 1.74	74.28 ± 1.37	**77.67 ± 1.81**	41.32 ± 1.38
Pendigits	90.22 ± 0.69	**93.94 ± 2.35**	81.42 ± 1.93	78.41 ± 1.44
Rank	1.57	1.57	**1.42**	3.0

[2] Multi-class joint boosting is a relatively new multi-class approach where instead of training independent classifiers for each object class, they are jointly trained. This training is performed by finding common features that can be shared across classes, leading to a robust feature extraction and a good generalization of the recognition problem.

Fig. 7. Triangular traffic sign classes.

Landmarks technique. The landmarks are located at the three corners of the signs. In order to learn each landmark at size 21×21 pixels (Fig. 1) we use cascades of classifiers. We used stratified 10-fold cross-validation to train each cascade of 10 levels and 500 negatives samples, with an expected error of 0.3. The correlograms used have a diameter of 150 pixels, 20 radius regions and 13 geometric circles with an enlarging factor of 1.3. As a result, we obtain a total of 780 features for each landmark correlogram including the object attributes and spatial positions.

The whole process of detection and recognition is illustrated for some test images in Fig. 9. First column corresponds to the detected landmark candidates labeled by color. The second column of the figure shows the combination of landmarks obtained by the highest likelihood of the contextual descriptors classifiers. At the third column one can see the final recognition results – the recovered object from the traffic sign database.

Table 4
Hit ratio results for the traffic sign database

Strategy	Detection accuracy
Fergus (Fergus and Zisserman, 2003)	94.3 ± 0.70
Boosting Context (Amores et al., 2005)	97.9 ± 0.60
Boosted Landmarks in Contextual Descriptors	99.1 ± 0.40

Table 4 shows the results on detecting traffic signs from the set of real samples by the three detectors. Our results for the test set are upon 99%. Up to our opinion it is very positive taking into account the high variability in appearance of the test signs (Fig. 8). Note that our technique is also robust to slight affine image transformations, illumination changes, and partial occlusion due to the descriptors used. Next step consists in extracting the region of the sign in order to classify it. Hence, we use the pixel-based features from the Boosted Landmarks detected objects to categorize among sign classes.

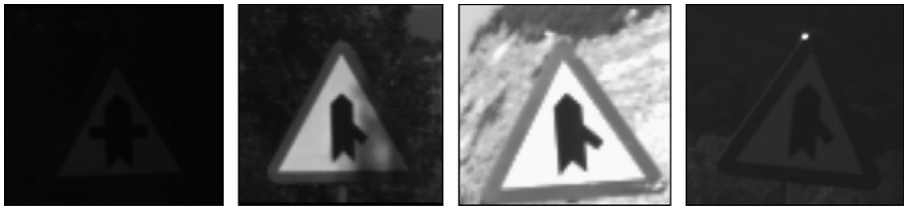


Fig. 8. Real triangular sign images in non-controlled conditions.

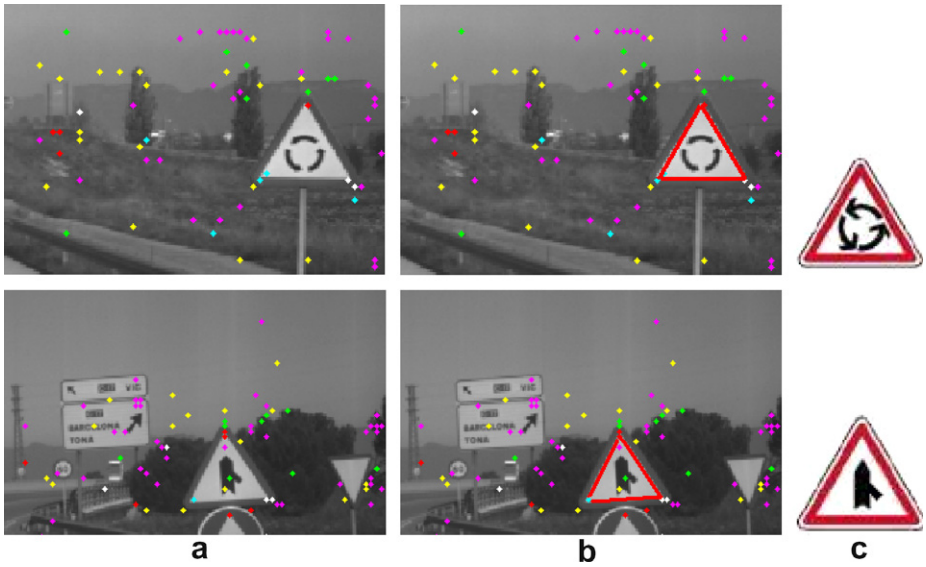
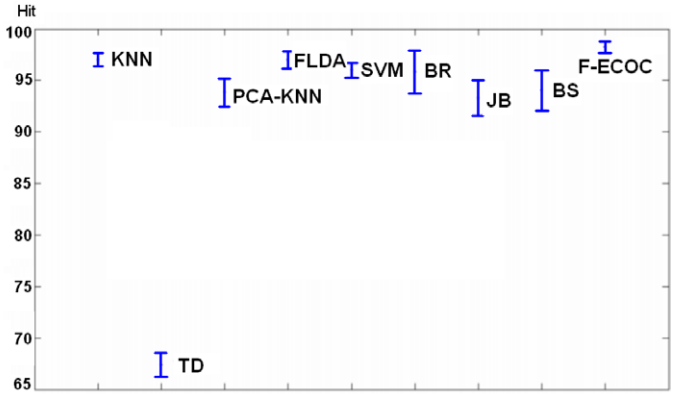


Fig. 9. Two examples of the whole procedure for real traffic sign images. (a) Landmark candidates for test images. (b) Predominant likelihoods of landmark combination. (c) Classification results (landmarks candidates are shown in color) (for interpretation of colour in this figure, the reader is referred to the web version of this article).

Method	Triangular
KNN	97.00±0.62
TD	67.36±1.22
PCAKNN	93.42±1.42
all pairs FLDA	96.88±0.78
SVM	96.11±0.71
BR	95.73±2.13
JB	93.37±1.68
BS	94.12±1.89
FOREST-ECOC	**98.25±0.53**

a



b

Fig. 10. (a) Recognition rate and confidence interval for the traffic sign database. (b) Graphical results.

For the recognition of the signs, we compared our proposed technique with the following classical classifiers: 3-Nearest neighbors (KNN), Tangent Distance (TD) (Simard et al., 1998) with invariant tangent vector respect translation, rotation and scaling, 99.9% of PCA followed by 3-Nearest neighbors (PCA-KNN) (Dudoit et al., 2002), all pairs FLDA with a previous 99.9% of PCA (Dudoit et al., 2002), Support Vector Machine with projection kernel Radial Basis Function and parameter gamma set to 1 (SVM) (Hsu et al., 2002), 40 runs of Gentle Adaboost with decision stumps using rectangular features on the integral image (BR) (Lienhart and Maydt, 2002; Kim et al., 2000), 40 runs of multi-class Joint Boosting with decision stumps (JB) (Murphy et al., 2003), 40 runs of Gentle Adaboost (Friedman et al., 1998) Sampling with FLDA and a previous 99.9% of PCA (BS), and our Forest-ECOC approach. Our technique is tested with $T = 2$ multiple trees using 40 runs of Gentle Adaboost with decision stumps to train the tree nodes classifiers to code and the Euclidean distance to decode. All the tests use 10-fold cross-validation and a two-tailed t-test to estimate the confidence interval. Fig. 10a shows the results obtained. We can observe that our system obtains the best result when compared to the different state-of-art classification techniques, and its compact confidence interval assure us to obtain stable results. Besides, the novel scheme processes a medium resolution image of 800×600 pixels in less than one second. These classification results are also graphically shown at Fig. 10b.

5. Conclusions

In this article, we introduced a novel, fast, and robust strategy for object detection and classification based on boosted contextual landmarks to detect and capture objects in cluttered scenes learning simultaneously the most relevant object features and their relations. Boosting is the base classifier and acts as feature selector, parts detector and recognizer. We show its accuracy on the Caltech database and solve a real traffic sign problem, comparing to well-known object recognition approaches. The procedure is invariant to small variations in scale, translation, global illumination, partial occlusions and to small affine transformations. Moreover, we presented a novel recognition technique called Forest-ECOC based on the embedding of multiple optimal trees in an ECOC framework. Our approach is not bounded by the number of classifiers, instead it allows constructing an ensemble of tree structures until the necessary performance is achieved. We validate this method using the UCI repository datasets and real traffic sign images obtaining very promising results, competing with an extended set of ten state-of-the-art multiclass recognition techniques. To improve the accuracy of our context of Boosted Landmarks, we are focusing on generating deformable contextual descriptors in order to allow to find elastic objects with higher deformations and to detect them from different points of view.

References

Agarwal, S., Awan, A., Roth, D., 2004. Learning to detect objects in images via a sparse. Transactions on PAMI 26 (11), 1475–1490.

Allwein, E., Schapire, R., Singer, Y., 2002. Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research, 113–141.

Amores, J., Sebe, N., Radeva, P., 2005. Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors. In: CVPR, vol. 2, pp. 769–774.

Baro, X., Vitria, J., 2004. Traffic sign detection on greyscale images. In: CCIA, pp. 209–216.

Belongie, S., Malik, J., Puzicha, J., 2000. Shape context: A new descriptor for shape matching and object recognition. In: NIPS, pp. 831–837.

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. Transactions in PAMI, 509–522.

Dietterich, T., Bakiri, G., 1991. Error-correcting output codes: A general method for improving multiclass inductive learning programs. In: Press, A. (Ed.), Ninth National Conference on Artificial Intelligence, pp. 572–577.

Dietterich, T., Bakiri, G., 2005. Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research 2, 263–286.

Dietterich, T., Kong, E., 1995. Error-correcting output codes corrects bias and variance. In: Prieditis, S., Russell, S. (Eds.), Proceedings of the 21th International Conference on Machine Learning, pp. 313–321.

Dudoit, S., Fridlyand, J., Speed, T., 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. In: JASA, pp. 77–87.

Escalera, S., Radeva, P., 2004. Fast greyscale road sign model matching and recognition. In: Recent Advances in Artificial Intelligence Research and Development, pp. 69–76.

Fergus, P.P.R., Zisserman, A., 2003. Object class recognition by unsupervised scale-invariant learning. In: CVPR.

Friedman, T., Hastie, T., Tibshirani, R., 1998. Additive logistic regression: A statistical view of boosting. The Annals of Statistics 38 (2), 337–374.

Hastie, T., Tibshirani, R., 1998. Classification by pairwise grouping. The Annals of Statistics 26 (5), 451–471.

Hong, P., Huang, T.S., 2004. Spatial pattern discovery by learning a probabilistic parametric model from multiple attributed relations graphs. In: International Workshop on Combinational Image Analysis, vol. 139, pp. 113–135.

Hsu, C., Chang, C., Lin, C.-J., 2002. A practical guide to support vector classification.

http://www.vision.caltech.edu/html-files/archive.html.

Kim, Y., Hahn, S., Zhang, B., 2000. Text filtering by boosting naïve Bayes classifiers. In: SIGIR Conference on Research and Development, pp. 168–175.

Lienhart, R., Maydt, J., 2002. An extended set of haar-like features for rapid object detection, pp. 155–162.

Murphy, K., Torralba, A., Freeman, W.T., 2003. Using the forest to see the trees: A graphical model relating features, objects, and scenes. In: Press, M. (Ed.), Advances in NIPS.

Pujol, O., Radeva, P., Vitrià, J., 2006. Discriminant ecoc: A heuristic method for application dependent design of error correcting output codes. Transactions on PAMI 28 (6), 1001–1007.

Ricci, F., Aha, D., 1998. Error-correcting output codes for local learners. European Conference on Machine Learning 1398, 280–291.

Schneiderman, H., 2004. Learning a restricted Bayessian network for object detection. In: CVPR.

Simard, P., LeCum, Y., Denker, J., Victorri, B., 1998. Transformation invariance in pattern recognition, tangent distance and tangent propagation. In: Orr, G., Muller, K. (Eds.), Neural Networks: Tricks of the Trade, pp. 239–274.

Torralba, A., Murphy, K., Freeman, W., 2004. Sharing visual features for multiclass and multiview object detection. In: CVPR, vol. 2, pp. 762–769.