Forest Extension of Error Correcting Output Codes and Boosted Landmarks

Sergio Escalera
CVC, Computer Science Department, UAB
Campus UAB, 08193 Bellaterra, Spain
sescalera@cvc.uab.es

Oriol Pujol
Dept. Matemàtica Aplicada i Anàlisi
UB, Gran Via 585, 08007, Barcelona, Spain
oriol@cvc.uab.es

Petia Radeva
CVC, Computer Science Department, UAB
Campus UAB, 08193 Bellaterra, Spain
petia@cvc.uab.es

Abstract

In this paper, we introduce a robust novel approach for detecting objects category in cluttered scenes by generating boosted contextual descriptors of landmarks. In particular, our method avoids the need of image segmentation, being at the same time invariant to scale, global illumination, occlusions and to small affine transformations. Once detected the object category, we address the problem of multiclass recognition where a battery of classifiers is trained able to capture the shared properties between the object descriptors across classes. A natural way to address the multiclass problem is using the Error Correcting Output Codes technique. We extend the ECOC technique proposing a methodology to construct a forest of decision trees that are included in the ECOC framework. We present very promising results on standard databases: UCI database and Caltech database as well as in a real image problem.

1. Introduction

Usually, the problem of object recognition (e.g. person identification) needs a previous addressing the category detection (e.g. face location). According to the way objects are described, three main families of approaches can be considered [7]: part-based, patch-based and region-based methods. Part-based approaches consider that an object is defined as a specific spatial arrangement of the object parts. An unsupervised statistical learning of constellation of parts and spatial relations is used in [5]. In [2] a representation integrating Boosting with constellations of contextual descriptors is defined, where the feature vector includes the bins that correspond to the different positions of the correlograms determining the object properties. Patch-based methods 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 [11], where objects are described by the best features obtained using 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. In [2], the selection of feature points is based on image contour points.

Once the object is located, it should be recognized using a kind of classification technique (support vector machines, nearest neighbor approach, linear discriminant analysis, etc.). Recently, Torralba et.al. [11] proposed a novel multiclass approach where instead of training independent classifiers for each object class, detectors for each class are jointly trained that leads to more robust object features chosen by the learners and better generalization of the recognition approach. Following the multitask framework, where a set of classifiers should learn in a natural way the features shared between categories, we choose to use the Error Correcting Output Codes (ECOC) [4] technique that has been shown to be a very successful multiclass framework due to its ability to extend any binary classifier to the multiclass classification problem. 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 [8]. In this paper, 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 matrix.

Our goal in this paper is two-fold: first, we introduce a novel approach for the detection of objects in cluttered scenes. On one hand, we use Boosted Landmarks to identify landmark candidates in the image without need to segment it. 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. Second, a new multiclass learning technique is introduced based on embedding a forest of optimal trees in an ECOC framework that allows to share features (tree nodes, base classifiers or dichotomies) in a very robust way.

2. Object Detection by Boosted Landmarks and Contexts

In this section we introduce an object detection method based on training the best discriminant features of the object.

2.1. Patch-based: Boosting landmarks

In order to avoid considering all possible ROIs of an image where an object can be located, we first have to find the possible locations of the object of interest helped by a set of landmarks. These landmarks are trained by means of a boosting procedure avoiding in this way the need of image segmentation. In particular, Gentle Adaboost [6] has been used since it has been shown to outperform most of the other boosting variants in real applications. This procedure is fed with the result of the rectangular features estimated on the Integral Image [3] over each landmark. This procedure is introduced in a cascade of weak classifiers [4], where each level of the cascade is specialized on a complex set of features to reject false positives. For example, in the case of triangular traffic signs we consider the six representative landmarks showed in fig. 1. From a training set of positive samples and a negative set of background images, we train each landmark in a cascade of weak classifiers. The presented scheme is invariant to scale, global illumination and to small object affine transformations.

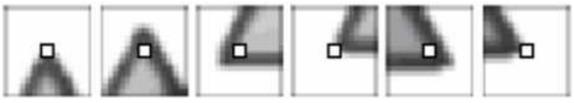

(a) (b) (c) (d) (e) (f)

Figure 1. Selected landmarks for triangular signs.

2.2. Parts-based: Contextual Descriptors

This step focuses on the spatial relationship among the previously detected landmarks. This approach is an extension of [2] in which a set of points of interest, $P = \{p_i\}_{i=1}^{N}$, where N is the number of interest points, coming from the edges of the image are used to build a constellation of multiscale correlograms. However, opposed to the work presented in [2], we use the information provided by landmarks instead of a set of contour points. Given n landmarks and their sets of detected candidates $L^1 = L_1^1 ... L_{i_j}^1, ..., L^n = L_1^n ... L_{i_n}^n$, where i_j is the number of instances for landmark j, for each combination of possible landmarks candidates $\{L_{j_i}^1, ..., L_{j_n}^n | j_1 \in \{1, ..., i_1\}, ..., j_n \in \{1, ..., i_n\}\}$, we generate n correlograms centered in the n chosen candidates that combined form a constellation. From this constellation we design a contextual descriptor vector $D = [D_1, , D_n]$, where D_i is the descriptor [9] vector associated to landmark i. Hence, the spatial relationship vector is obtained as the values of the correlogram bins for each of the landmarks. For example, using the 6 landmarks shown in fig. 1, the spatial descriptor vector is 6 × N bins in length. Using Gentle Adaboost, we train at the same time the relevant features, which in our case are the detected landmarks, and their spatial relations. As additional information we use the contour points of the image fig. 2(c) shows an example of a correlogram at a detected landmark. In fig. 2(d) we observe the detected object, which contextual descriptor, defined by the combination of detected landmarks, has been accepted as positive using the classifier trained by boosting.

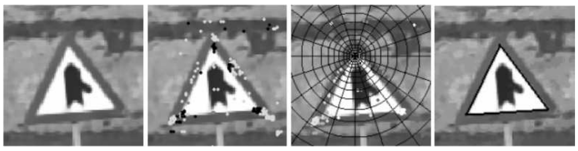
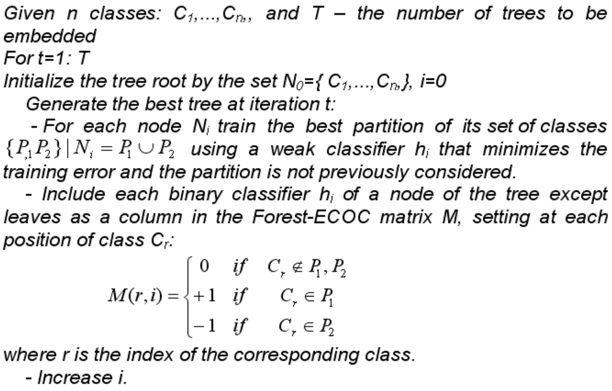

(a) (b) (c) (d)

Figure 2. (a) Input image. (b) Labeled landmarks. (c) Contextual descriptors. (d) Detected sign.

3. Object Recognition by Forest-ECOC

Once located the object category (e.g. a traffic sign) we proceed with the object recognition following a multi-task framework. ECOC were born as a general framework for handling multiclass problems, sharing information across classes to improve its accuracy. The basis of this framework is to create a codeword for each of N_c classes (up to N_c codewords). Arranging the codewords as rows of a matrix, the "coding matrix" M is defined, where $M \in \{-1, 0, 1\}^{N_c \times n}$, and n is the code length. From encoding point of view of learning, the matrix M can be seen as n binary learning problems corresponding to the n columns of the matrix, (coded by +1, 0 and -1 according to the class membership). A zero value indicates that a particular class is not relevant for a given binary classifier. As a result of the outputs of the n dichotomies, a code is obtained for each data point in the test set, that comparing with the base codewords of M (corresponding to the matrix rows), is assigned to the class with the "closest" codeword.

In [8] a method for embedding a tree structure in the ECOC framework is proposed. Taking this work as a baseline, we propose the use of multiple trees embedding forming a Forest-ECOC. However, opposed to the discriminant tree proposed in [8], we use the classification score to create each node of the tree. The tree with a maximal-score at each node is called "optimal" tree. Beginning from a root containing all classes, first, we build a binary tree where each node shows the best partition of classes that minimizes the training error. This process of finding the partition of classes set is done recursively until getting sets of single classes corresponding to the tree leaves. In the case that we consider the best T partitions, it allows us to create multiple trees. These trees are embedded in the ECOC matrix forming the Forest-ECOC, to get an ensemble of trees. This algorithm is shown in fig. 3.



```
Given n classes: $C_1,...,C_n$, and T – the number of trees to be embedded
For t=1: T
Initialize the tree root by the set $N_0$={ $C_1,...,C_n$}, i=0
    Generate the best tree at iteration t:
     - For each node $N_i$ train the best partition of its set of classes $\{P_1 P_2\} | N_i = P_1 \cup P_2$ using a weak classifier $h_i$ that minimizes the training error and the partition is not previously considered.
    - Include each binary classifier $h_i$ of a node of the tree except leaves as a column in the Forest-ECOC matrix M, setting at each position of class $C_r$:
        $M(r,i) = \begin{cases} 0 & if \quad C_r \notin P_1, P_2 \\ +1 & if \quad C_r \in P_1 \\ -1 & if \quad C_r \in P_2 \end{cases}$
where r is the index of the corresponding class.
    - Increase i.
```

Figure 3. Training algorithm for Forest-ECOC.

For a given multiclass problem, generally, a set of two or three optimal trees gives accurate enough results. An example of two optimal-trees and the Forest-ECOC matrix for a toy problem are shown in fig. 4 and 5, respectively. Given an input image to test with the Forest-ECOC matrix, we generate the Forest-ECOC vector where each position is the result of the binary classifier of the columns of the matrix.

The classification is done assigning the label of the class with minimal distance between the row codeword and the test codeword using the standard decoding based on the Hamming Distance Estimation of $d_j(x, y^j) = \sum_{i=1}^{n} |(x_i - y_i^j)|/2$, where d_j is the distance to the row j, n is the number of dichotomies, and $x = (x_1, x_2, ..., x_n)$ and $y = (y_1^j, y_2^j, ..., y_n^j)$ are the results of classification of a test example and base codeword of class j, respectively.

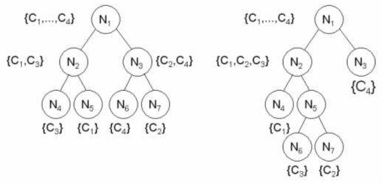
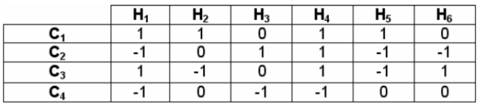


Figure 4. Two optimal trees for a toy problem.

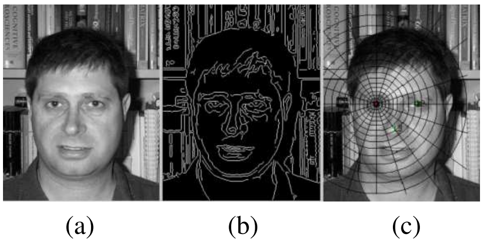


	H_1	H_2	H_3	H_4	H_5	H_6
C_1	1	1	0	1	1	0
C_2	-1	0	1	1	-1	-1
C_3	1	-1	0	1	-1	1
C_4	-1	0	-1	-1	0	0

Figure 5. Forest-ECOC matrix for a toy problem, where H_1, H_2 and H_3 correspond to classifiers of N_1, N_2 and N_3 from the first tree of fig. 4, and H_4, H_5 and H_6 to N_1', N_2' and N_5' from the second tree.

4. Results

Given both (detection and recognition) parts of our approach, we split the validation in two steps: validation of the Boosted Landmarks of Contextual Descriptors and the validation of the recognition approach based on the Forest-ECOC. In order to compare the accuracy of our detector, we tested it on the Caltech database [1] considering the following objects: car side, faces and motorbike, training only three landmarks from the models of each database. In fig. 6, the models, contour points, landmarks trained, and one correlogram for face database are shown. To test the method we used 20% of samples to train landmarks and contextual descriptors by boosting, and the rest to test. Table 1 shows that our results surpass those proposed in [2] and [5]. In the Motorbike database, the detected landmarks are less discriminant, so our procedure decreases in this case.



(a) (b) (c)

Figure 6. Fergus database. (a) Model. (b) Contour map. (c) Landmark correlogram.

We have also tested the false alarm rate using the background set of images from the Caltech database, obtaining just one false positive from the 500 instances. We apply our

Category	Fergus [5]	Boosting Context [2]	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%

Table 1. Hit ratio for the Fergus database.

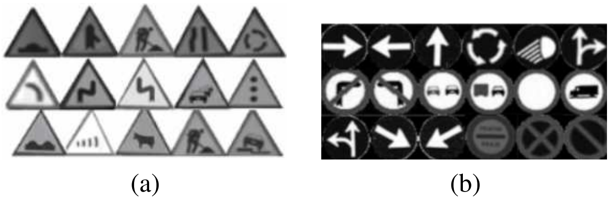
UCI	JB	all pairs FLDA	Forest ECOC
Dermathology	97.00±0.44	97.10±0.36	**98.01±0.45**
Ecoli	**86.40±0.27**	85.60±0.25	**86.40±0.24**
Balance	**90.00±0.40**	88.50±0.80	**89.20±0.70**
Iris	95.20±0.80	**97.40±0.70**	**97.40±0.70**

Table 2. Results for UCI databases.

detector to solve a real traffic sign detection and recognition problem. We used a database of 300 traffic sign images obtained by a Geovan in non-controlled outdoor conditions, where 200 signs have been used to train (fig. 7). The cascades to learn each landmark of fig. 1 use all rectangular features trained at size 21 × 21 pixels. Each cascade has 10 levels of 100 positives samples and 100 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 of factor 1.3, having a total of 780 features for each landmark correlogram including the object attributes and spatial positions. Our result for a test set of 100 samples has been of 99% hit ratio compared to the best results of 92% achieved in [2].

In order to validate our Forest-ECOC classification technique we tested it on the UCI repository datasets. We compare the proposed method with two very high performance classification techniques: Joint Boosting (JB) [11] and all-pairs ECOC [10] technique. All the tests are calculated using ten-fold cross-validation and a two-tailed t-test for the confidence interval. Observing the results in table 2 we can conclude that Forest-ECOC compares favourably with the other approaches and is a promising technique for multi-class recognition.

We applied the method to the recognition of the located triangular and circular traffic signs (fig. 7). Table 3 shows the results, where we can observe that our novel multiclass recognition approach is very competitive, achieving the best results when compared with the other techniques.



(a) (b)

Figure 7. Group of triangular (a) and circular (b) traffic signs.

Traffic signs	JB	1-vs-1 ECOC	Forest-ECOC
Triangular	86.36±1.98	95.57±1.21	**97.65±0.84**
Circular	90.97±0.97	94.35±1.13	**97.73±0.94**

Table 3. Results for traffic signs.

5. Acknowledgements

This work was supported in part by the projects TIC2003-00654, FIS-G03/1085, FIS-PI031488, and MI-1509/2005.

6. Conclusions

We introduced a fast and robust novel framework to detect and capture objects in cluttered scenes. The procedure is invariant to scale, global illumination, occlusion and to small affine transformations. We show its accuracy on the Caltech database and solve a real traffic sign problem, comparing with well-known detection approaches. Moreover, we presented a novel recognition technique called Forest-ECOC based on the embedding of multiple optimal trees in an ECOC framework. We validated this method using the UCI repository datasets and real traffic sign images obtaining very promising results, competing with state-of-art multiclass recognition techniques.

References

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

[2] J. Amores, N. Sebe, and P. Radeva. Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors. In *CVPR (in press)*, 2005.

[3] X. Baro and J. Vitria. Traffic sign detection on greyscale images. In *CCIA*, pages 209–216, 2004.

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

[5] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, 2003.

[6] T. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 38(2):337–374, 1998.

[7] K. Murphy, A.Torralba, and W.T.Freeman. Using the forest to see the trees: A graphical model relating features, objects, and scenes. In M. Press, editor, *Advances in NIPS*, 2003.

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

[9] S.Belongie, J.Malik, , and J.Puzicha. Shape matching and object recognition using shape contexts. *Transactions in PAMI*, pages 509–522, April 2002.

[10] T.Hastie and R. Tibshirani. Classification by pairwise grouping. *The annals of statistics*, 26(5):451–471, 1998.

[11] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. In *CVPR*, volume 2, pages 762–769, 2004.