

Circular Blurred Shape Model for Multi-class Symbol Recognition

^{1,2}Sergio Escalera, ^{1,3}Alicia Fornés, ^{1,2}Oriol Pujol, ^{1,3}Josep Lladós, and ^{1,2}Petia Radeva

Abstract—In this paper, we propose a Circular Blurred Shape Model descriptor to deal with the problem of symbol detection and classification as a particular case of object recognition. Feature extraction is performed by capturing the spatial arrangement of significant object characteristics in a correlogram structure. Shape information from objects is shared among correlogram regions, where a prior blurring degree defines the level of distortion allowed in the symbol, making the descriptor tolerant to irregular deformations. Moreover, the descriptor is rotation invariant by definition. We validate the effectiveness of the proposed descriptor in both multi-class symbol recognition and symbol detection domains. In order to perform symbol detection the descriptors are learnt using a cascade of classifiers. In the case of multi-class categorization, the new feature space is learnt using a set of binary classifiers which are embedded in an Error-Correcting Output Codes design. The results over four symbol data sets show significant improvements of the proposed descriptor compared to the state-of-the-art descriptors. In particular, the results are even more significant in those cases where symbols suffer from elastic deformations.

Index Terms—Symbol recognition, Multi-class categorization, Object detection, Symbol description, Error-Correcting Output Codes.

I. INTRODUCTION

Object recognition can be divided in two main problems: object detection and object categorization. Object detection techniques must be able to locate the target object while discarding most part of the image, meanwhile multi-class categorization must classify the object by its corresponding true class given a large set of possible classes. Symbol recognition is a particular problem of object recognition. Symbols are graphical entities made by humans to be read by humans. The problem of symbol recognition is a classical interest among the community of Document Image Analysis and Recognition. Recognition of technical documents or logo spotting for document database retrieval are typical applications. In the last years, Symbol Recognition has been also focused on images of natural scenes (e.g. traffic signs). Rotation, partial occlusions, elastic deformations, intra-class and inter-class variations, or high variability among symbols due to different writing styles (in the case of handwritten documents), are just a few problems in this domain.

Shape is one of the most important visual cues for describing objects, and as well as color or texture, it is widely used

for describing the content of the object. There is an increasing interest in the development of good shape recognition methods in the area of Pattern Recognition. In general, the design of a shape-based approach can be divided in two main steps: the definition of expressive and compact shape descriptors, and the formulation of robust classification methods for detection and classification.

Shape representation is a difficult task because of several object distortions, such as occlusions, elastic deformations, discontinuities, or noise. A good shape descriptor should guarantee inter-class compactness and intra-class separability, even when describing noisy and distorted shapes. The main techniques for shape recognition are reviewed in [1]. They are mainly classified in continuous and structural approaches. Zernike moments and Angular Radial Transform are examples of continuous approaches, which extract information from the whole shape region. Zernike moments [2] maintain properties of the shape, and are invariant to rotation, scale, and deformations. Angular Radial Transform (ART) [3] decomposes the shape in an orthogonal basis, making use of a radial and angular function. It has good performance for general shapes and uses few features by descriptor. On the contrary, other continuous approaches only use the external contour (silhouette) for computing the features, i.e. Curvature Scale Space (CSS) or Shape context [4]. CSS [5] is a standard of the MPEG7 [6] that is tolerant to rotation, but it can only be used for closed curves. Shape Context [4] can work with non-closed curves, and has good performance in hand drawn symbols, because it is tolerant to deformations, but it requires point-to-point alignment of the symbols.

Structural approaches use to represent shapes with relational information between compounding primitives. Straight lines and arcs are usually the basic primitives, which approximate contours and skeletons. Strings, graphs or trees represent the relations between these primitives. The similarity measure is performed by string, tree, or graph matching. Attributed graph grammars, Deformable models, and Region Adjacency Graphs are a few examples of structural approaches. Attributed graph grammars [7] can cope with repetitive subpatterns, while Region Adjacency Graphs [8] reach good performance in front of distortions in hand drawn documents. Deformable models on graph based representations of vectorized line drawings [9] are invariant to distortions and rotation, but require good initialization and robust edge detection.

Symbol descriptors robust to some affine transformations and occlusions are not effective enough when dealing with elastic deformations. Thus, the research of a descriptor that can cope with elastic deformations and non-uniform distortions is

¹Centre de Visió per Computador, Campus UAB, Edifici O, 08193 Bellaterra, Barcelona, Spain.

²Dept. Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007, Barcelona, Spain.

³Computer Science Dept., Campus UAB, 08193, Bellaterra, Barcelona, Spain.

{sescalera,afornes,oriol,josep,petia}@cvc.uab.es

still required. In the work of [10], the Blurred Shape Model (BSM) was presented. It is a descriptor that can deal with soft, rigid, and elastic deformations, but it is sensitive to rotation.

In this paper, we present an evolution of the Blurred Shape Model descriptor, which not only copes with distortions and noise, but it is also rotation invariant. The Circular Blurred Shape Model (CBSM) codifies the spatial arrangement of object characteristics using a correlogram structure. Based on a prior blurring degree, object characteristics are shared among correlogram regions. By rotating the correlogram so that the major descriptor densities are aligned to the x -axis, the descriptor becomes rotation invariant. We validate the descriptor in two scenarios: symbol detection and categorization. In order to deal with the problem of symbol detection [11], different pattern recognition methods are proposed in the literature, such as geometric features, region-based approaches using connected components, or structural symbol representation [12]. In our case, the new descriptor is learnt using a cascade of classifiers with Adaboost, and tested with a windowing strategy in order to locate the target object. The validation of the detection procedure is performed over an architectural image data set and an old music score image data set. In this case, our method shows better performance than the standard SIFT descriptor by tolerating large changes in symbol orientations. Moreover, the original BSM descriptor requires the object alignment previous to its description, which considerably increases the computational cost in comparison to the proposed circular approach.

Referring the categorization of several object classes, many classification techniques have been developed. One of the most well-known techniques is the Adaboost algorithm, which has been shown to be suitable for feature selection and achieves high performance when applied to binary categorization tasks [13]. The extension of this approach to the multi-class case is usually solved by combining the binary classifiers in a voting procedure, such as one-versus-one or one-versus-all voting schemes. In order to extend binary classifiers to the multi-class case, Dietterich et. al. [13] proposed the Error Correcting Output Codes framework (ECOC), which benefits from error correction properties, obtaining successful results [14]. In this paper, we learn the CBSM features by a dichotomizer based on the Adaboost classifier, and then, we combine the binary problems in an ECOC configuration, which extends the system to deal with multi-class categorization problems. The multi-class classification methodology has been used to compare the state-of-the-art descriptors: BSM, Zernike, Zoning and SIFT on the public MPEG7 data set and a grey-level symbol data set.

The paper is organized as follows: Section 2 presents the Circular Blurred-Shape Model descriptor. Section 3 shows the multi-class categorization and object detection methodologies considered to evaluate the CBSM descriptor. Section 4 presents the experimental evaluation on different binary and grey-level multi-class symbol categorization and detection problems. Finally, concluding remarks and perspectives are presented.

II. CIRCULAR BLURRED-SHAPE MODEL

In this section, we present a circular formulation of the Blurred Shape Model descriptor [10]. By defining a correlogram structure from the center of the object region, spatial arrangement of object parts is shared among regions defined by circles and sections. The method aims to achieve a rotation invariant description by rotating the correlogram according to the predominant region density, which implies the full redefinition of the BSM descriptor. We divide the description of the algorithm into three main steps: the definition of the correlogram parameters, the descriptor computation, and the rotation invariant procedure.

Correlogram definition: Given a number of concentric circles C , a radius R , a number of sections S , and an image region I , a correlogram $B = \{b_{\{1,1\}}, \dots, b_{\{C,S\}}\}$ is defined as a radial distribution of sub-regions of the image, as shown in Figures 1(a),(b). Each region b is defined by its centroid coordinates b^* (see Fig. 1(c)). Then, the regions around b are defined as the neighbors of b . Note that depending on the spatial location of the analyzed region, different number of neighbors can be defined (see Fig. 1(d)). Different correlogram structures are shown in Figure 2 for different values of C and S .

Descriptor computation: In order to compute the CBSM descriptor, first, a pre-processing of the input region I to obtain the shape features is required. For several symbols, relevant shape information can be obtained by means of a contour map (though based on the object properties we can define a different pre-processing step). In this paper, we use the Canny edge detector procedure.

Given the object contour map, each point of the image belonging to a contour is taken into account in the description process (see Fig. 1(e)). First of all, the distances from the contour point \mathbf{x} to the centroids of its corresponding region and neighboring regions are computed. The inverse of these distances is normalized by the sum of total distances. These values are then added to the corresponding positions of the descriptor vector ν (see Fig. 1(f)). This makes the description tolerant to irregular deformations. Concerning the computational complexity, note that for a correlogram of $C \times S$ sectors and k contour points considered for obtaining the CBSM descriptor, only $O(k)$ simple operations are required. The description procedure is detailed in Algorithm 1.

Rotation invariant descriptor: In order to obtain a rotation invariant description, a second step is included in the description process. We look for the main diagonal G_i of the correlogram B that maximizes the sum of the descriptor values. This diagonal is then taken as a reference for rotating the descriptor. The orientation in the rotation process, so that G_i is aligned to the x -axis, is the one corresponding to the highest density of the descriptor at both sides of G_i . This procedure is detailed in Algorithm 2. A visual result of the rotation invariant process can be observed in Figure 3, in which two bats with different descriptor orientations are rotated and aligned.

In this way, the output descriptor ν for an input region I represents a distribution of probabilities of the symbol structure

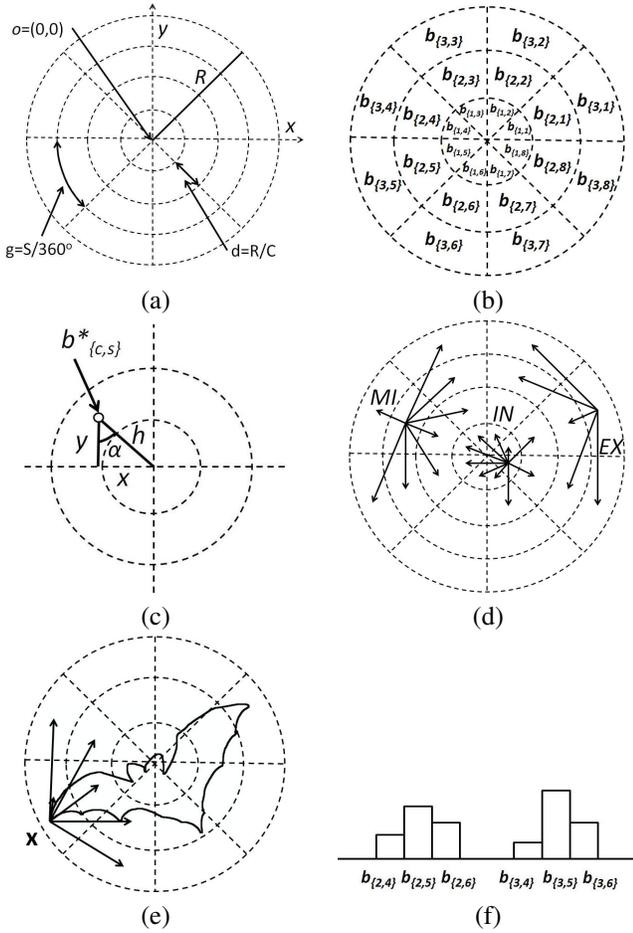


Fig. 1. (a) CBSM correlative parameters, (b) regions distribution, (c) region centroid definition, (d) region neighbors, (e) object point analysis, and (f) descriptor vector update after the analysis of point x .

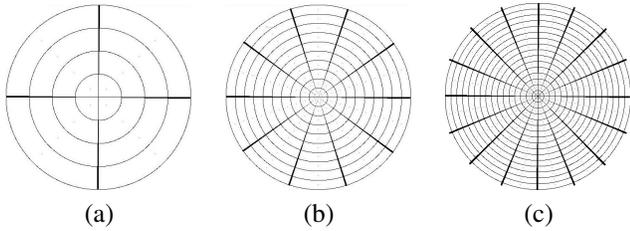


Fig. 2. Correlative structures obtained for different $C \times S$ sizes: (a) 4×4 , (b) 10×10 , and (c) 16×16 .

considering spatial distortions, where the number of regions (defined by parameters C and S) defines the blurring degree allowed. The blurring degree defines the degree of spatial information taken into account in the description process. In Figure 3, a bat instance from the public MPEG7 data set [15] is described with different $C \times S$ correlative sizes. Note that when we increase the number of regions, the description becomes more local. Thus, optimal parameters of C and S should be obtained for each particular problem (e.g. via cross-validation, splitting the training data into two subsets, one to train and the remaining one to validate the method parameters). The selected number of regions (and consequently, the blurring degree) is the one which attains the highest performance on

Algorithm 1 Circular Blurred Shape Model Description Algorithm.

Require: a binary image I (of dimensions $Y \times Z$), a number of concentric circles C , and a number of sections S

Ensure: descriptor vector ν , and the set of bins B

- 1: **Define** $R = \max(Y/2, Z/2)$ as the radius of the most outer concentric circle.
- 2: **Define** $d = R/C$ and $g = S/360$ as the distance between consecutive concentric circles and the degrees between consecutive sectors, respectively (Figure 1(a)).
- 3: **Define** $B = \{b_{\{1,1\}}, \dots, b_{\{C,S\}}\}$ as the set of bins for the circular description of I , where $b_{c,s}$ is the bin of B between distances $[(c-1)d, c \cdot d)$ to the origin of coordinates o and x -axis (Figure 1(b)).
- 4: **Define** $b_{\{c,s\}}^* = (\sin \alpha d, \cos \alpha d)$ the centroid coordinates of bin $b_{\{c,s\}}$, and $B^* = \{b_{\{1,1\}}^*, \dots, b_{\{C,S\}}^*\}$ the set of centroids in B (Figure 1(c)).
- 5: **Define** $X_{b_{\{c,s\}}^*} = \{b_1, \dots, b_{c,s}\}$ as the sorted set of the elements in B^* so that $d(b_{\{c,s\}}^*, b_i^*) \leq d(b_{\{c,s\}}^*, b_j^*)$, $i < j$.
- 6: **Define** $N(b_{\{c,s\}})$ as the neighbor regions of $b_{\{c,s\}}$, defined by the initial elements of $X_{b_{\{c,s\}}^*}$:

$$N(b_{\{c,s\}}) = \begin{cases} X', |X'| = S + 3 & \text{if } b_{\{c,s\}} \in IN \\ X', |X'| = 9 & \text{if } b_{\{c,s\}} \in MI \\ X', |X'| = 6 & \text{if } b_{\{c,s\}} \in EX \end{cases}$$

being IN , MI , and EX , the inner, middle, and outer regions of B , respectively (Figure 1(d)).

- 7: **Initialize** $\nu_i = 0$, $i \in [1, \dots, C \cdot S]$, where the order of indices in ν are:
- 8: $\nu = \{b_{\{1,1\}}, \dots, b_{\{1,S\}}, b_{\{2,1\}}, \dots, b_{\{2,S\}}, \dots, b_{\{C,1\}}, \dots, b_{\{C,S\}}\}$
- 9: **for** each point $x \in I$, $I(x) = 1$ (Figure 1(e)) **do**
- 10: $D = 0$
- 11: **for** each $b_i \in N(b_x)$ **do**
- 12: $d_i = d(x, b_i) = \|x - b_i^*\|^2$
- 13: $D = D + \frac{1}{d_i}$
- 14: **end for**
- 15: Update the probabilities vector ν positions as follows (Figure 1(f)):
- 16: $\nu(b_i) = \nu(b_i) + \frac{1}{d_i D}$, $\forall i \in [1, \dots, C \cdot S]$
- 17: **end for**
- 18: **Normalize** the vector ν as follows:
- 19: $d' = \sum_{i=1}^{C \cdot S} \nu_i$, $\nu_i = \frac{\nu_i}{d'}$, $\forall i \in [1, \dots, C \cdot S]$

Algorithm 2 Rotation invariant ν description.

Require: a number of circles C , a number of sections S and a set of bins B

Ensure: Rotation invariant descriptor vector ν^k

- 1: **Define** $G = \{G_1, \dots, G_{S/2}\}$ the $S/2$ diagonals of B , where $G_i = \{\nu(b_{\{1,i\}}), \dots, \nu(b_{\{C,i\}}), \dots, \nu(b_{\{1,i+S/2\}}), \dots, \nu(b_{\{C,i+S/2\}})\}$
- 2: Select G_i so that $\sum_{j=1}^{2C} G_i(j) \geq \sum_{j=1}^{2C} G_k(j)$, $\forall k \in [1, \dots, S/2]$
- 3: **Define** L_G and R_G as the left and right areas of the selected G_i as follows:
- 4: $L_G = \sum_{j,k} \nu(b_{\{j,k\}})$, $j \in [1, \dots, C]$, $k \in [i+1, \dots, i+S/2-1]$
- 5: $R_G = \sum_{j,k} \nu(b_{\{j,k\}})$, $j \in [1, \dots, C]$, $k \in [i+S/2+1, \dots, i+S-1]$
- 6:
- 7: **if** $L_G > R_G$ **then**
- 8: B is rotated $k = i + S/2 - 1$ positions to the left:
- 9: $\nu^k = \{\nu(b_{\{1,k+1\}}), \dots, \nu(b_{\{1,S\}}), \nu(b_{\{1,1\}}), \dots, \nu(b_{\{1,k\}}), \dots, \nu(b_{\{C,k+1\}}), \dots, \nu(b_{\{C,S\}}), \nu(b_{\{C,1\}}), \dots, \nu(b_{\{C,k\}})\}$
- 10: **else**
- 11: B is rotated $k = i - 1$ positions to the right:
- 12: $\nu^k = \{\nu(b_{\{1,S\}}), \dots, \nu(b_{\{1,S-k+1\}}), \nu(b_{\{1,1\}}), \dots, \nu(b_{\{1,S-k\}}), \dots, \nu(b_{\{C,S\}}), \dots, \nu(b_{\{C,S-k+1\}}), \nu(b_{\{C,1\}}), \dots, \nu(b_{\{C,S-k\}})\}$
- 13: **end if**

the validation subset, defining the optimum number of sizes, encoding the different distortions on each particular problem, and offering the required trade-off between inter-class and intra-class variabilities in a problem-dependent way.

The CBSM correlative is defined by means of a number of sectors S and number of concentric circles C in a linear correlative design. It implies that the area of the outer sectors is higher than the area corresponding to inner sectors. Since we define the same importance to all analyzed shape points, it seems intuitive to define sectors with the same area. However, in this paper we define a linear concentric circles definition which implies more local description on the center of the

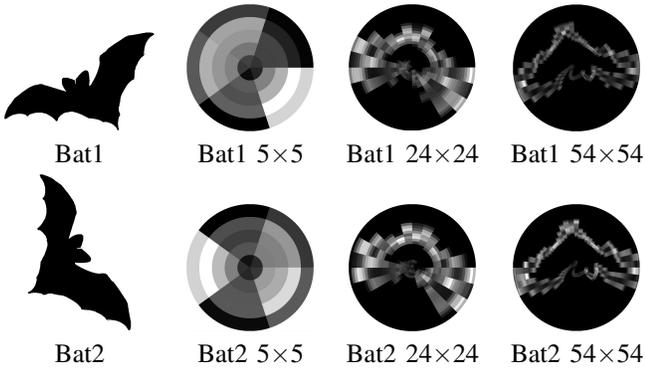


Fig. 3. Examples of image descriptors at different sizes for two object instances. The more regions are used for the description, the more local information about the shape is obtained. Notice that the two descriptors are correctly rotated and aligned.

description meanwhile the distortion degree allowed at the external sectors is increased. We use this approximation based on the fact that the outer appearance of symbols is usually higher compared to the inner variabilities (i.e. the external strokes in hand-drawn symbols). On the other hand, if we want to define a correlogram structure where all sectors have the same area, we simply need to change the distance among correlogram sectors to satisfy the new constraints.

III. CBSM DETECTION AND CLASSIFICATION SYSTEM

For the sake of completeness, in this section we overview the object categorization and symbol detection methodologies considered for validating the proposed descriptor.

A. Symbol classification system

The proposed symbol classification system consists of two different stages: description and classification. For the first stage, the above described rotation invariant CBSM descriptor is computed. For the second stage, the Error Correcting Output Codes (ECOC) framework is used. The whole process is shown in Figure 4.

Error Correcting Output Codes [13] is a meta-learning strategy that divides a multi-class problem in a set of binary problems, solves them individually, and aggregates their responses into a final multi-class framework. ECOC have been successfully applied to many machine vision tasks [16], [17], [18], [19], showing interesting properties in statistical learning, reducing both the bias and the variance of the base classifiers [20].

The ECOC meta-learning algorithm consists of two steps: the learning/coding step, where an ECOC encoding matrix is constructed in order to define the combination of classifiers in the coding matrix \mathbf{T} , and the decoding step, where a new sample \mathbf{x} is classified according to the set of binary classifiers of \mathbf{T} . Formally, given a set of N training samples $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, where each \mathbf{x}_i belongs to a class $C_i \in \{C_1, \dots, C_K\}$, an ECOC encoding consists of constructing M binary problems using the original K classes. Each binary problem splits two meta-classes, and values +1

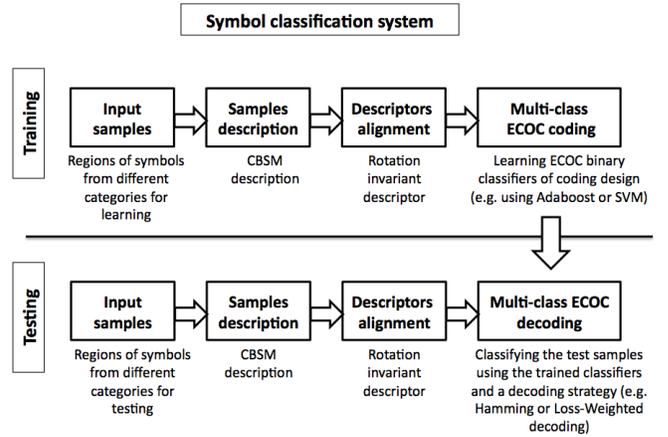


Fig. 4. Symbol Classification System. In the training step, the CBSM descriptor is computed for all the symbols, and the ECOC encoding matrix is constructed for defining the combination of classifiers. In the testing step, the CBSM descriptor is computed for the input symbols, and after their alignment, they are classified using the ECOC decoding algorithm.

and -1 are assigned to each class belonging to the first and second meta-classes, respectively. If a class does not belong to any meta-class, the membership value is set to 0. This creates a $K \times M$ matrix \mathbf{T} . When a new sample must be classified, the outputs of the classifiers trained on each binary problem (columns of the matrix \mathbf{T}) are used to construct the codeword that is compared with each row of the matrix \mathbf{T} . The class codeword with the minimum distance is selected as the classifier output. The ECOC scheme allows to represent in a common framework well-known strategies such as one-versus-all or all-pairs (one-versus-one), as well as more sophisticated problem-dependent encodings, namely discriminant ECOC [21] or sub-class ECOC [14], without a significant increment of the codeword length. Literature shows that one of the most straightforward and well-performing approaches disregarding the properties of the particular base learner is the *one-versus-one* strategy.

The final part of the ECOC process is based on defining a suitable distance for comparing the output of the classifiers with the base codewords. The authors of [22] have recently shown that *weighted decoding* achieves the minimum error with respect to most state-of-the-art decoding measures. The weighted decoding strategy decomposes the decoding step of the ECOC technique in two parts: a weighting factor for each code position and any general decoding strategy. In [22] the authors have shown that for obtaining a successful decoding, two conditions must be fulfilled: the bias induced by the zero symbol should be zero and the dynamic range of the decoding strategy must be constant for all the codewords. The complete decoding strategy weights the contribution of the decoding at each position of the codeword by the elements of a weighting matrix \mathbf{W} that ensures that both conditions are fulfilled. As such, the final decoding strategy is defined as,

$$\delta(y, \mathbf{T}(i, \cdot)) = \sum_{j=1}^M \mathbf{W}(i, j) \cdot \mathcal{L}(\mathbf{T}(i, j) \cdot h_j(x))$$

where

$$w(i, j) = \begin{cases} r_i(S, \mathbf{T}(\cdot, j), \mathbf{T}(i, j)) & \mathbf{T}(i, j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{j=1}^M w(i, j) = 1, \forall i \in \{1 \dots K\}$$

We define the meta-class relative accuracy (r -value) of class k on the set S given the definition of the meta-class ρ as,

$$r_k(S, \rho, i) = \frac{\#\text{elements of class } k \text{ classified as meta-class } i \text{ in the set } S}{\#\text{elements belonging to class } k \text{ in the set } S} \quad (1)$$

where ρ defines which classes belong to which meta-class.

The second part of the weighting decoding relies in a base decoding strategy. In this paper, we use the Linear Loss-based decoding as base decoding strategy. Linear Loss-based decoding was introduced by Allwein et al. [23] and is defined in the following way: given the input sample \mathbf{x} and the binary code y resulted of applying all the dichotomizers (h_1, h_2, \dots, h_M) to the input test sample, the decoding value is defined as follows,

$$\delta(y, \mathbf{T}(i, \cdot)) = \sum_{j=1}^M \mathcal{L}(\mathbf{T}(i, j) \cdot h_j(x))$$

where $\mathbf{T}(i, \cdot)$ denotes the codeword for class i , $h_j(x)$ is the prediction value for dichotomizer j , and \mathcal{L} is a loss function that represents the penalty due to the miss-classification. In the case of Linear Loss-based decoding, we have $\mathcal{L}(\rho) = -\rho$.

Note that the ECOC framework just requires $K \cdot M$ tests to perform multi-class classification, being K the number of possible object categories and M the number of trained classifiers.

B. Symbol detection system

In order to design a symbol detection methodology, two stages must be defined. A first stage (namely training) should learn to distinguish among the target object and *background* (i.e. learning a binary classifier). A second stage (namely testing) should perform a search over the whole image using the trained classifier in order to locate those regions containing the target object. The whole process is shown in Figure 5.

For the first step, we propose to learn a binary classifier using Adaboost [24] with a set of positive and negative object instances. Since we need to apply this classifier to a huge number of regions in the second step, the final detection time for an image is very high. In order to address this limitation, Viola & Jones introduced a cascade architecture of multiple *strong classifiers* [25]. The underlying idea is to use only the necessary computation cost in order to reject a non-object regions while more complex analysis is performed in unclear cases. Those regions that arrive to the last stage of the cascade are classified as objects, and then selected as object regions, meanwhile the rest of the regions are rejected. Each stage of the cascade only analyzes the objects accepted by the previous stages, and thus, the non-objects are analyzed only until they are rejected by a stage. The number of applied classifiers is

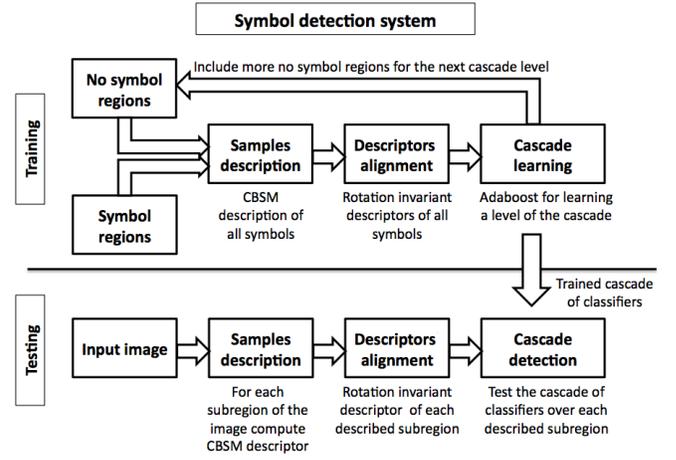


Fig. 5. Symbol Detection System. In the training step, the CBSM descriptor is computed for all the symbols, and the cascade of classifiers is used for learning the positive and negative object instances. In the testing step, the CBSM descriptor is computed for all the candidate subregions, and the cascade of classifiers is used for detecting the regions containing the target object.

Algorithm 3 Attentional cascade training algorithm.

Require: A set of positive examples P and a set of negative examples N , a maximum false alarm rate f , a minimum accuracy a , and a number of cascade levels L .

Ensure: A cascade of strong classifiers h .

- 1: **for** $i = 1$ to L **do**
- 2: $F_i \leftarrow 1, n_i \leftarrow 0$
- 3: **while** $F_i > f$ **do**
- 4: $n_i \leftarrow n_i + 1$
- 5: Use P and N to train a classifier with n_i features using Adaboost
- 6: $F_i \leftarrow$ Evaluate current cascaded classifier on validation set
- 7: Decrease threshold for the i th classifier until the current cascaded classifier satisfies a detection rate of a (this also affects F_i)
- 8: **end while**
- 9: $N \leftarrow 0$
- 10: Evaluate the current cascaded detector on the set of non-object images and put any false detections into the set N .

11: **end for**

reduced exponentially due to the cascade architecture. This strategy is detailed in Algorithm 3.

Once the cascade of classifiers is learnt, a windowing strategy is applied on the whole test image. The method is described in Algorithm 4.

Algorithm 4 Object detection using a cascade of classifiers.

Require: An image I , a cascade of classifiers h , an initial window size S_I , a final window size S_F , a shift s , and an increment i .

Ensure: Target object regions R

- 1: **for** windows W of size S_I , increasing by i , to S_F **do**
- 2: **for** each region r in I of size W with shift s among regions, increasing by i **do**
- 3: test cascade h over region r
- 4:
$$h(r) = \begin{cases} 1 & \text{if detected as positive (object instance)} \\ & \text{save region } \rightarrow R = R \cup r \\ 0 & \text{if detected as negative (background).} \end{cases}$$
- 5: **end for**
- 6: **end for**

IV. EXPERIMENTAL EVALUATION

We divide the experimental evaluation in two main blocks: multi-class symbol categorization and symbol detection.

A. Multi-class symbol categorization

In order to present the multi-class categorization results, we discuss the data, methods, and validation of the experiments:

- *Data*: For comparing our CBSM multi-class methodology, we used two multi-class data sets: first, the public 70-class MPEG7¹ binary repository data set [15], which contains a high number of classes with different appearance of symbols from a same class, including rotation. The second data set is a 17-class data set of grey-level symbols², which contains the common distortions from real-environments, such as illumination changes, partial occlusions, or changes in the point of view.

- *Methods*: The descriptors considered in the comparison results are SIFT [26], BSM [10], Zoning [1], and Zernike moments [2]. The details of the descriptors used for the comparison results are the following: the optimum correlogram size of the CBSM descriptor is estimated applying cross-validation over the training set using a 10% of the samples to validate the different sizes of $S = \{8, 12, 16, 20, 24, 28, 32\}$ and $C = \{8, 12, 16, 20, 24, 28, 32\}$. For the sake of fairness, the Zoning and BSM descriptors are set to the same number of regions as the CBSM descriptor. Rotation invariance for the BSM descriptor is achieved by means of principal components alignment before descriptor computation [10]. Concerning the Zernike moments descriptor, 7 moments are used. Gentle Adaboost with 50 decision stumps [24] is used for training the binary problems of the one-versus-one ECOC design [23] with the Loss-Weighted decoding (LW) [22] to solve the multi-class categorization problems. We also consider a Support Vector Machine with a Radial Basis Function base classifier for the ECOC design with $C = 1$ and $\gamma = 1$ and a 3-Nearest Neighbor classifier in the comparison results. The regularization parameter C and the γ parameter are set to one for the experiments. We selected this parameter after a preliminary set of evaluations. We decided to keep the parameter fixed for the sake of simplicity and easiness of replication of the experiments, though we are aware that this parameter might not be optimal for the analyzed data sets.

- *Validation*: The classification score is computed by means of stratified ten-fold cross-validation [27], testing for the 95% of the confidence interval CI with a two-tailed t-test [28], computed as:

$$CI = \frac{1.96 \sigma_{X_j}}{\sqrt{NT}} \quad (2)$$

where σ_{X_j} is the standard deviation of the performance of the tests X_j , and NT is the number of tests.

Next, we describe the experiments performed, comparing our descriptor with state-of-the-art descriptors over two multi-class categorization problems (with binary and grey-level symbols).

1) *MPEG7 Multi-classification data set*: In this experiment, we used the 70 object categories from the public MPEG7 binary object data set [15] to compare the whole set of

¹MPEG7 Repository Database: <http://www.cis.temple.edu/~latecki/research.html>

²These data sets and ground truths are publicly available under request to the authors of this paper.

Descriptor	3NN	ECOC LW Adaboost	ECOC LW SVM
CBSM	71.84(6.73)	80.36(7.01)	78.32(6.38)
BSM	65.79(8.03)	77.93(7.25)	78.14(8.12)
Zernike	43.64(7.66)	51.29(5.48)	49.33(6.37)
Zoning	58.64(10.97)	65.50(6.64)	61.22(6.87)
SIFT	29.14(5.68)	32.57(4.04)	28.18(5.91)

TABLE I
CLASSIFICATION ACCURACY AND CONFIDENCE INTERVAL (IN BRACKETS) ON THE 70 MPEG7 SYMBOL CATEGORIES FOR THE DIFFERENT DESCRIPTORS USING 3-NEAREST NEIGHBOR AND THE ONE-VERSUS-ONE ECOC SCHEME WITH GENTLE ADABOOST AND RBF SVM AS THE BASE CLASSIFIERS.

descriptors in a multi-class categorization problem. A pair of samples of some classes of this data set are shown in Figure 6.

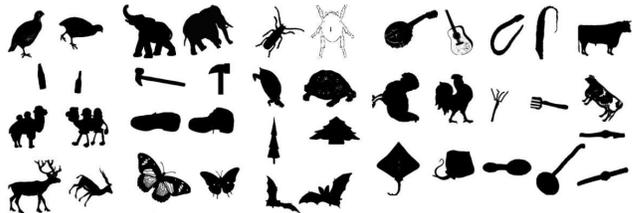


Fig. 6. MPEG7 samples.

The classification results and confidence interval after testing using a stratified ten-fold cross-validation with a 3-NN and the ECOC configuration with Gentle Adaboost and RBF SVM base classifiers are shown in Table I. The value in brackets corresponds to the confidence interval. Note that the best performance is obtained by the CBSM descriptor for all classifiers, followed in all cases by the BSM descriptor. Moreover, the ECOC configurations always obtain higher performance than classifying with a nearest neighbor classifier. On the other hand, Adaboost performs better than RBF SVM as ECOC base classifier in this data set.

2) *Grey-scale Multi-classification symbol data set*: The second data set of symbols consists of grey-level samples from 17 different classes, with a total of 550 samples acquired with a digital camera from real environments. The samples are taken so that there are large affine transformations, partial occlusions, background influence, and high illumination changes. A pair of samples for each of the 17 classes are shown in Fig. 7. Some examples of the data set of this experiment and their corresponding CBSM descriptors are shown in Fig. 8. In this type of data sets, the SIFT descriptor has shown to be the one which attains the highest performance in comparison to the state-of-the-art descriptors. Due to this reason, we compare our CBSM with the SIFT descriptor [26] as well as with the BSM descriptor [10].

Table II shows the performances and confidence intervals obtained in this experiment using a ten-fold cross-validation with the CBSM, BSM, and SIFT descriptors in an one-versus-one ECOC scheme with Gentle Adaboost as the base classifier and Loss-Weighted decoding. One can see that the result obtained by the CBSM descriptor adapted to grey-scale symbols outperforms the result obtained by the SIFT and BSM descriptors. This difference is produced in this data set because

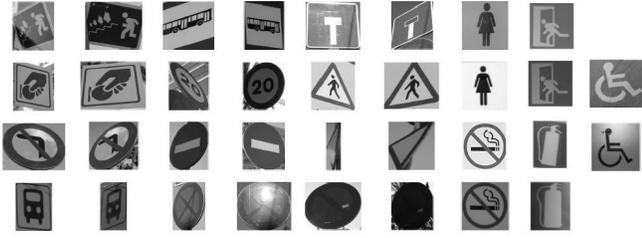


Fig. 7. Grey-scale symbol data set samples.

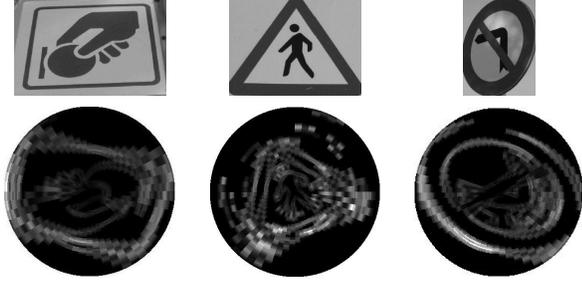


Fig. 8. CBSM descriptors from samples of the grey-level symbols data set.

of the high changes in the point of view of the symbols and the background influence, which produce significant changes of the SIFT orientations. Moreover, the rotation invariance of the CBSM descriptor makes it faster and more robust than the BSM descriptor with previous alignment based on principal components.

B. Symbol detection

In order to show the evaluation of the detection results, we first describe the test data, the methods that have been compared with our algorithm, and validation framework to measure the experimental evaluation.

- *Data*: To test the detection CBSM methodology, we selected the predefined architectural plan files of the Smart Draw software [29] and the old handwritten musical scores from a collection of modern and old handwritten musical scores (19th century) of the Archive of the Seminar of Canet de Mar, Barcelona³.

- *Methods*: The descriptors considered in the comparison results are SIFT [26] and BSM [10]. The parameters used are the same than in the previous experiment. We trained 10 levels of the cascade with Gentle Adaboost classifier with 50 decision stumps [24], and 5000 random background images from Google were used as negative set.

³These data sets and ground truths are publicly available under request to the authors of this paper.

CBSM	BSM	SIFT
77.82(6.45)	75.23(7.18)	62.12(9.08)

TABLE II

CLASSIFICATION ACCURACY AND CONFIDENCE INTERVAL OF THE CBSM, BSM, AND SIFT DESCRIPTORS ON THE GREY-SCALE SYMBOLS DATA SET USING AN ONE-VERSUS-ONE ECOC SCHEME WITH GENTLE ADABOOST AS THE BASE CLASSIFIER.

- *Validation*: We apply the evaluation framework of [30] for the detection rate criterion. The detection rate measures how correct the detector selects the target regions, which have been previously manually labeled. Then, the accuracy is measured by the amount of overlapping between the detected region and the labeled one. We consider that two regions are matched if they satisfy:

$$1 - \frac{R_d \cap R_o}{R_d \cup R_o} < \epsilon \quad (3)$$

where R_d is the detected region and R_o is the original one. We set the maximum overlap error ϵ to 40%, as in [30]. Moreover, we introduce the false alarm rate criterion, defined as the ratio between the number of detected regions that do not match which the original labeled ones (false positives) and the total number of detected regions. This measure should be as small as possible.

Next, we describe the experiments performed, comparing our descriptor with state-of-the-art descriptors on two binary and grey-level symbol detection problems.

- 1) *Symbol detection in raster images of scanned architectural plans*: In this experiment, we used 20 predefined architectural plan files of the Smart Draw software [29]. We trained a cascade of classifiers with the parameters previously defined for the CBSM, BSM, and SIFT descriptors. We used 30 positive door symbol samples for training the cascade. Since there will be many overlapped detections, we will define an accepted positive region as the region which has a minimum of 3 positive detections with an intersection area greater than the 70% of the area of the smallest overlapped detection. Note that many positive windows can appear around the target object. In this way, we also discard false positive isolated detection. Two examples of doors and their CBSM rotation invariant descriptors are shown in Figure 9.

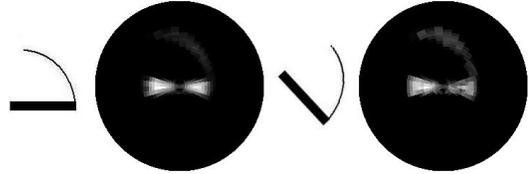


Fig. 9. Two examples of door positive images and their corresponding CBSM visual descriptors.

Some visual results testing the CBSM detection procedure with a window shift of five pixels (which has been experimentally set) are shown in Figure 11. Note that all the doors are detected even when connected with different types of walls and on different rotation degrees. The numerical detection results for three descriptors are shown in Figure 10(a). From the total number of doors in the 20 architectural plan images, the 32 test doors were successfully detected by the three descriptors using the measure of eq. (3), obtaining a hit ratio of 100%. Moreover, only one false positive region was detected in the case of the CBSM descriptor, corresponding to a 3% of the detected regions. Note that one positive region from the thousands of analyzed regions is insignificant⁴.

⁴A video file showing the learning and symbol detection process for architectural symbols has been submitted together with the paper.

	Objects detected	False alarm
CBSM	100%	3%
SIFT	100%	9%
BSM	100%	18%

(a)

	Objects detected	False alarm
CBSM	93.33%	18.92%
SIFT	70%	45.59%
BSM	83.33%	38.78%

(b)

Fig. 10. (a) Detection results over the architectural plan images. (b) Detection results over the musical score images.

2) Symbol detection in old handwritten musical scores:

In this last experiment, we used 20 old handwritten musical scores from a collection of modern and old handwritten musical scores (19th century) of the Archive of the Seminar of Canet de Mar, Barcelona. We trained a cascade of classifiers with the parameters previously defined for the CBSM, BSM, and SIFT descriptors. We compare with the SIFT descriptor since it is the most widely applied on grey-level intensity images. We used 144 positive music clef samples for training the cascade.

As in the previous experiment, we consider a positive region if there are a minimum of three intersections, and discard false positive isolate detection. Some results testing the CBSM detection procedure with a window shift of also five pixels on different staves are shown in Figure 12. Note that all the clefs are detected. One false positive is shown at the end of the music sheet. Notice that under this false positive a rotation of the region appears so that it looks as the beginning of a staff, where a clef can appear. It is the main reason why the detection procedure confuses the region. The numerical detection results for the three descriptors are shown in Figure 10(b). In this case, the degradation of the images reduces the accuracy of three descriptors in comparison to the previous case. In particular, from the total number of 30 test clefs in the images, the best accuracy is obtained by the CBSM descriptor, detecting 28 symbols using the measure of eq. (3), which corresponds to a hit ratio of 93.33%. Regarding the false positives, the lowest false alarm rate is also obtained by the CBSM descriptor, detecting only 7 false positive regions.

V. CONCLUSIONS AND PERSPECTIVES

In this last section, we summarize the contributions of our work and present open issues.

A. Conclusions

In this paper, a Circular Blurred Shape Model descriptor has been presented. The new descriptor is suitable to describe and recognize, in a fast way, symbols that can suffer from several distortions, such as occlusions, rigid or elastic deformations, discontinuities, or noise. The descriptor encodes the spatial arrangement of symbol characteristics using a correlogram structure. A prior blurring degree defines the level of degradation allowed to the symbol. Moreover, the descriptor correlogram is rotated guided by the major density, becoming rotation invariant.

The new descriptor is used to solve object detection and multi-class categorization problems. In the case of multi-class

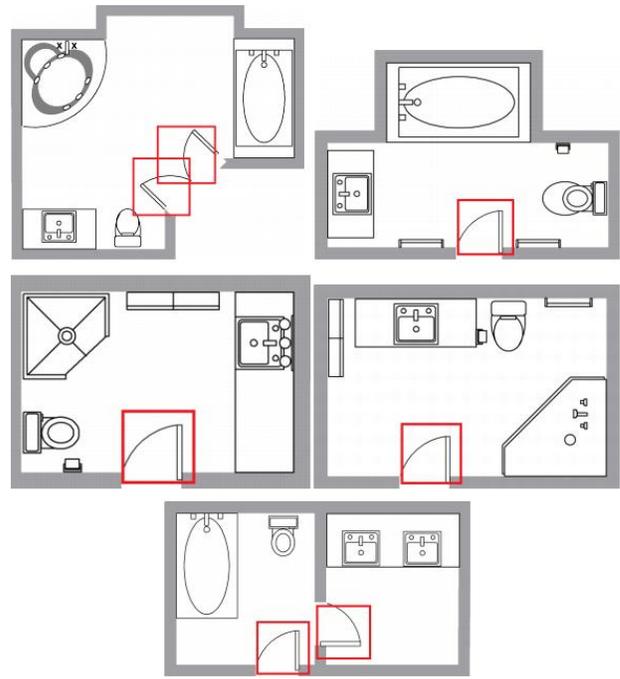


Fig. 11. Smartdraw architectural plan images and door symbol detection.



Fig. 12. Clef detection in old handwritten music score images. A false positive is shown at the bottom of the figure.

symbol recognition, the new symbol descriptions are learnt using Adaboost binary classifiers and embedded in an Error-Correcting Output Codes framework. The experimental results on different binary and grey-level multi-class categorization problems show that the CBSM descriptor obtains higher performance than the state-of-the-art descriptors, specially when classifying high number of symbol classes that suffer from irregular deformations.

For the detection problem, the descriptor is learnt using

a cascade of classifiers with Adaboost to discard non-object regions, and tested over whole images, detecting the target objects. The symbol detection procedure presented in this paper has shown to robustly locate object instances in documents, such as binary symbols in architectural plans and grey-level symbols in old handwritten musical scores, outperforming the accuracy of the state-of-the-art descriptors and reducing the false alarm rate.

B. Perspectives

Contour map image points have been used in this work. However, depending on the kind of objects to be described, different types of features could be considered and blurred among the CBSM sectors. In this sense, the contours could be labeled based on different structure properties (such as those defined in [31]) and then the CBSM descriptor could be defined of this new set of features.

ACKNOWLEDGMENT

This work has been partially supported by projects TIN2009-14633-C03-03, TIN2008-04998, FIS PI061290, and CONSOLIDER-INGENIO CSD 2007-00018.

REFERENCES

- [1] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.
- [2] A. Khotanzad and Y. Hong, "Invariant image recognition by zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.
- [3] W. Kim and Y. Kim, "A New Region-based Shape Descriptor," Hanyang University and Konan Technology, Tech. Rep., 1999.
- [4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
- [5] F. Mokhtarian and A. Mackworth, "Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 34–43, 1986.
- [6] B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley, 2002.
- [7] H. Bunke, "Attributed Programmed Graph Grammars and their Application to Schematic Diagram Interpretation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, no. 6, pp. 574–582, 1982.
- [8] J. Lladós, E. Martí, and J. J. Villanueva, "Symbol recognition by error-tolerant subgraph matching between region adjacency graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1137–1143, 2001.
- [9] E. Valveny and E. Martí, "Hand-drawn symbol recognition in graphic documents using deformable template matching and a bayesian framework," in *Proc. of the 15th International Conference on Pattern Recognition*, vol. 2, 2000, pp. 239–242.
- [10] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, , and J. Lladós, "Blurred shape model for binary and grey-level symbol recognition," *Pattern Recognition Letters*, vol. 30, pp. 1424–1433, 2009.
- [11] K. Tombre, S. Tabbone, and P. Dosch, "Musings on Symbol Recognition," *Lecture Notes in Computer Science*, vol. 3926, pp. 23–34, 2006.
- [12] D. Zuwala and S. Tabbone, "A Method for Symbol Spotting in Graphical Documents," *Lecture Notes in Computer Science*, vol. 3872, pp. 518–528, 2006.
- [13] T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.
- [14] S. Escalera, D. Tax, O. Pujol, P. Radeva, and R. Duin, "Subclass problem-dependent design of error-correcting output codes," *IEEE Transactions in Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1041–1054, 2008.
- [15] "Mpeg7 repository." [Online]. Available: <http://www.cis.temple.edu/latecki/research.html>
- [16] R. Ghani, "Combining labeled and unlabeled data for text classification with a large number of categories," in *Proc. of the IEEE International Conference on Data Mining*, vol. 2, no. 1, 2001, pp. 597–598.
- [17] T. Windeatt and G. Ardeshir, "Boosted ECOC ensembles for face recognition," in *Proc. of the International Conference on Visual Information Engineering*, 2003, pp. 165–168.
- [18] J. Kittler, R. Ghaderi, T. Windeatt, and J. Matas, "Face verification using error correcting output codes," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 755–760.
- [19] J. Zhou and C. Suen, "Unconstrained numeral pair recognition using enhanced error correcting output coding: a holistic approach," in *Proc. of the International Conference on Document Analysis and Recognition*, vol. 1, 2005, pp. 484–488.
- [20] T. Dietterich and E. Kong, "Error-correcting output codes corrects bias and variance," in *Proc. of the International Conference on Machine Learning*, 1995, pp. 313–321.
- [21] O. Pujol, P. Radeva, and J. Vitrià, "Discriminant ECOC: A heuristic method for application dependent design of error correcting output codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1001–1007, 2006.
- [22] S. Escalera, O. Pujol, and P. Radeva, "On the decoding process in ternary error-correcting output codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 120–134, 2010.
- [23] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal on Machine Learning Research*, vol. 1, pp. 113–141, 2002.
- [24] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, no. 2, pp. 337–374, 2000.
- [25] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2002.
- [26] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal on Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [27] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [28] K. P. Johnson, R., *Elementary Statistics*. Duxbury Press, 2006.
- [29] "Smartdraw software." [Online]. Available: <http://www.SmartDraw.com>
- [30] K. Mikolajczyk, T. Tuytelaars, and C. Schmid, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1–2, pp. 43–72, 2005.
- [31] V. Ferrari, F. Jurie, and C. Schmid, "Accurate object detection with deformable shape models learnt from images," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–7.



systems (<http://www.maia.ub.es/~sergio>).

Sergio Escalera received the B.S. and M.S. degrees from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2003 and 2005, respectively. He obtained the Ph.D. degree on Multi-class visual categorization systems at Computer Vision Center, UAB. He is a researcher at the Computer Vision Center and a Lecturer with the Department of Applied Mathematics and Analysis, Universitat de Barcelona. His research interests include, between others, machine learning, statistical pattern recognition, visual object recognition, and human computer interaction



Alicia Fornés received the B.S. degree from the Universitat de les Illes Balears (UIB) in 2003 and the M.S. degree from the Universitat Autònoma de Barcelona (UAB) in 2005. She obtained the Ph.D. degree on writer identification of old music scores from the UAB in 2009. She is currently a postdoctoral researcher in the UAB and the Computer Vision Center. Her research interests include document analysis, symbol recognition, historical documents, and writer identification.



Oriol Pujol received the Ph.D. degree in Computer Sciences from the Universitat Autònoma de Barcelona, Spain, in 2004. He is currently an Associate Professor with the Department of Applied Mathematics and Analysis, Universitat de Barcelona. His research interest includes statistical machine learning techniques for object recognition and medical imaging analysis.



Josep Lladós received the degree and Ph.D. in Computer Sciences in 1991 and 1997 from the Universitat Autònoma de Barcelona and the Université Paris, respectively. Currently he is an Associate Professor at the Computer Sciences Department of the Universitat Autònoma de Barcelona and a staff researcher of the Computer Vision Center, where he is also the director. His current research fields are document analysis, graphics recognition and structural and syntactic pattern recognition. J. Lladós is an active member of the IAPR. He was the recipient of the

IAPR-ICDAR Young Investigator Award in 2007. Josep Lladós also created the company ICAR Vision Systems, a spin-off of the Computer Vision Center working on Document Image Analysis, after winning the entrepreneurs award from the Catalonia Government on business projects on Information Society Technologies in 2000.



Petia Radeva received the Ph.D. degree from the Universitat Autònoma de Barcelona (UAB), Spain. Her Ph.D. dissertation was focused on the development of physics-based models applied to image analysis. She is currently a researcher at the Computer Vision Center and an Associate Professor with 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.