# Circular Blurred Shape Model for Multiclass Symbol Recognition

Sergio Escalera, Alicia Fornés, Oriol Pujol, Josep Lladós, and 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. The feature extraction is performed by capturing the spatial arrangement of significant object characteristics in a correlogram structure. The 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 the multiclass symbol recognition and symbol detection domains. In order to perform the symbol detection, the descriptors are learned using a cascade of classifiers. In the case of multiclass categorization, the new feature space is learned using a set of binary classifiers which are embedded in an error-correcting output code design. The results over four symbol data sets show the 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 the symbols suffer from elastic deformations.**

***Index Terms*—Error-correcting output codes, multiclass categorization, object detection, symbol description, symbol recognition.**

## I. Introduction

OBJECT RECOGNITION can be divided into two main problems: object detection and object categorization. The object detection techniques must be able to locate the target object while discarding most part of the image; meanwhile, the multiclass 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. The recognition of technical documents or logo spotting for document database retrieval is a typical application. In the last years, symbol recognition has also been focused on the images of natural scenes (e.g., traffic signs). The rotation, partial occlusions, elastic deformations, intraclass and interclass variations, and 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 into two main steps: the definition of expressive and compact shape descriptors and the formulation of robust classification methods for the 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 interclass compactness and intraclass separability, even when describing noisy and distorted shapes. The main techniques for shape recognition are reviewed in [1]. They are mainly classified into continuous and structural approaches. The Zernike moments and angular radial transform (ART) are examples of continuous approaches, which extract information from the whole shape region. The Zernike moments [2] maintain the properties of the shape and are invariant to the rotation, scale, and deformations. The angular radial transform [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 the descriptor. On the contrary, other continuous approaches only use the external contour (silhouette) for computing the features, i.e., the curvature scale space (CSS) or shape context [4]. The CSS [5] is a standard of the MPEG-7 [6] that is tolerant to rotation, but it can only be used for closed curves. The shape context [4] can work with nonclosed 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.

The structural approaches are used to represent the shapes with relational information between the compounding primitives. The straight lines and arcs are usually the basic primitives, which approximate the contours and skeletons. The strings, graphs, or trees represent the relations between these primitives. The similarity measure is performed by string, tree, or graph matching. The attributed graph grammars, deformable models, and region adjacency graphs are a few examples of the structural approaches. The attributed graph grammars [7] can cope with repetitive subpatterns while the region adjacency graphs

Manuscript received April 19, 2009; revised March 11, 2010; accepted July 11, 2010. This work was supported in part by projects TIN2009-14633-C03-03, TIN2008-04998, and CONSOLIDER-INGENIO CSD 2007-00018 from the Ministerio de Ciencia y Educacind FIS PI061290 from the Instituto de Salud Carlos III. This paper was recommended by Associate Editor E. Santos.

S. Escalera, O. Pujol, and P. Radeva are with the Centre de Visió per Computador, Campus Universitat Autònoma de Barcelona, Edifici O, 08193 Barcelona, Spain. They are also with the Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, 08007 Barcelona, Spain (e-mail: sescalera@cvc.uab.es; oriol@cvc.uab.es; petia@cvc.uab.es).

A. Fornés and J. Lladós are with the Centre de Visió per Computador, Campus Universitat Autònoma de Barcelona (UAB), Edifici O, 08193 Barcelona, Spain. They are also with the Computer Science Department, Campus UAB, 08193 Barcelona, Spain (e-mail: afornes@cvc.uab.es; josep@cvc.uab.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

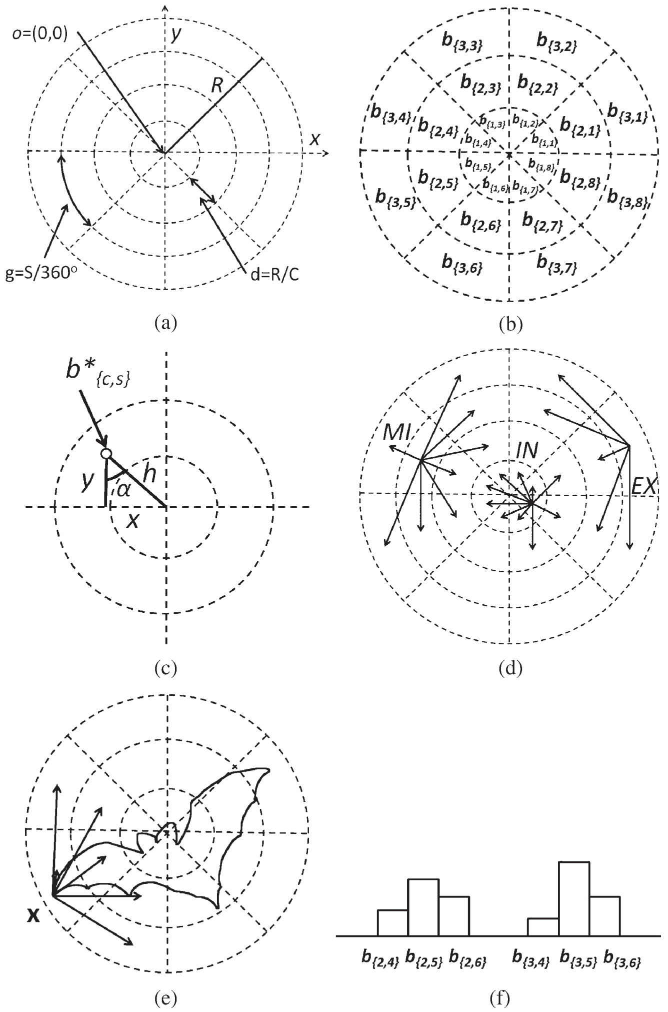
Digital Object Identifier 10.1109/TSMCB.2010.2060481

[8] reach good performance in front of distortions in the hand-drawn documents. The deformable models on the graph-based representations of vectorized line drawings [9] are invariant to distortions and rotation but require good initialization and robust edge detection.

The 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 nonuniform distortions is still required. In the work of Escalera *et al.* [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 BSM descriptor, which not only copes with distortions and noise but also is rotation invariant. The circular BSM (CBSM) codifies the spatial arrangement of the object characteristics using a correlogram structure. Based on a prior blurring degree, the 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 learned using a cascade of classifiers with Adaptive Boosting (AdaBoost) and tested with a windowing strategy in order to locate the target object. The validation of the detection procedure is performed over architectural and old-music-score image data sets. In this case, our method shows a better performance than the standard scale-invariant feature transform (SIFT) descriptor by tolerating large changes in the 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 to 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 multiclass case is usually solved by combining the binary classifiers in a voting procedure, such as the one-versus-one or one-versus-all voting schemes. In order to extend the binary classifiers to the multiclass case, Dietterich and Bakiri [13] proposed the error-correcting output code (ECOC) framework, which benefits from the 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 the multiclass categorization problems. The multiclass classification methodology has been used to compare the state-of-the-art descriptors–BSM, Zernike, Zoning, and SIFT–on the public MPEG-7 and grey-level-symbol data sets.



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

Fig. 1. (a) CBSM correlogram parameters. (b) Region distribution. (c) Region centroid definition. (d) Region neighbors. (e) Object point analysis. (f) Descriptor vector update after the analysis of point $x$.

This paper is organized as follows. Section II presents the CBSM descriptor. Section III shows the multiclass categorization and object detection methodologies considered to evaluate the CBSM descriptor. Section IV presents the experimental evaluation on different binary and grey-level multiclass symbol categorization and detection problems. Finally, the concluding remarks and perspectives are presented in Section V.

## II. CBSM

In this section, we present a circular formulation of the BSM descriptor [10]. By defining a correlogram structure from the center of the object region, the spatial arrangement of object parts is shared among the 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

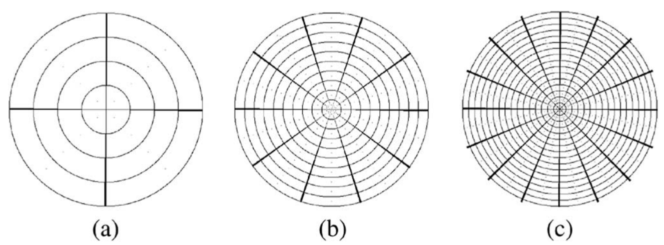


Fig. 2. Correlogram structures obtained for different $C \times S$ sizes. (a) 4 × 4. (b) 10 × 10. (c) 16 × 16.

distribution of the subregions of the image, as shown in Fig. 1(a) and (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 numbers of neighbors can be defined [see Fig. 1(d)]. Different correlogram structures are shown in Fig. 2 for different values of $C$ and $S$.

**Descriptor computation:** In order to compute the CBSM descriptor, first, a preprocessing of the input region $I$ to obtain the shape features is required. For several symbols, the relevant shape information can be obtained by means of a contour map (though based on the object properties, we can define a different preprocessing 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 the total distances. These values are then added to the corresponding positions of the descriptor vector $\nu$ [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 the $O(k)$ simple operations are required. The description procedure is detailed in Algorithm 1.

**Algorithm 1** CBSM description algorithm.

```
Require: a binary image $I$ (of dimensions $YxZ$), a number of concentric circles $C$, and a number of sections $S$.
Ensure: the descriptor vector $\nu$ and the set of bins $B$.
  Define $R = \max(Y/2, Z/2)$ as the radius of the most outer concentric circle.
  Define $d = R/C$ and $g = S/360$ as the distance between the consecutive concentric circles and the degrees between the consecutive sectors, respectively [see Fig. 1(a)].
  Define $B = \{b_{\{1,1\}}, \ldots, b_{\{C,S\}}\}$ 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 between interval angles $[(s-1)g, s \cdot g)$ to the origin of the coordinates $o$ and $x$-axis [see Fig. 1(b)].
  Define $b^*_{\{c,s\}} = (\sin\ \alpha\ d, \cos\ \alpha\ d)$ as the centroid coordinates of bin $b_{\{c,s\}}$ and $B^* = \{b^*_{\{1,1\}}, \ldots, b^*_{\{C,S\}}\}$ as the set of centroids in $B$ [see Fig. 1(c)].
  Define $X_{b_{\{c,s\}}} = \{b_1, \ldots, b_{c \cdot s}\}$ as the sorted set of the elements in $B^*$ so that $d(b^*_{\{c,s\}}, b^*_i) \leq d(b^*_{\{c,s\}}, b^*_j),\ i < j$.
  Define $N(b_{\{c,s\}})$ as the neighbor regions of $b_{\{c,s\}}$, defined by the initial elements of $X_{b_{\{c,s\}}}$
      $N(b_{\{c,s\}}) = \begin{cases} X', |X'| = S+3 & \text{if } b_{\{c,s\}} \in IN \\ X', |X'| = 9 & \text{if } b_{\{c,s\}} \in MI \\ X', |X'| = 6 & \text{if } b_{\{c,s\}} \in EX \end{cases}$
  where $IN$, $MI$, and $EX$ are the inner, middle, and outer regions of $B$, respectively [see Fig. 1(d)].
  Initialize $\nu_i = 0,\ i \in [1, \ldots, C \cdot S]$, where the order of indices in $\nu$ are as follows:
  $\nu = \{b_{\{1,1\}}, \ldots, b_{\{1,S\}}, b_{\{2,1\}}, \ldots, b_{\{2,S\}}, \ldots, b_{\{C,1\}}, \ldots, b_{\{C,S\}}\}$
  for each point $\mathbf{x} \in I$, $I(\mathbf{x}) = 1$ [see Fig. 1(e)] do
    $D = 0$
    for each $b_i \in N(b_{\mathbf{x}})$do
      $d_i = d(\mathbf{x}, b_i) = \|\mathbf{x} - b^*_i\|^2$
      $D = D + (1/d_i)$
    end for
    Update the probability vector $\nu$ positions as follows [see Fig. 1(f)]:
    $\nu(b_i) = \nu(b_i) + (1/d_i D), \forall i \in [1, \ldots, C \cdot S]$
  end for
  Normalize the vector $\nu$ as follows:
  $d' = \sum_{i=1}^{C \cdot S} \nu_i,\ \nu_i = \nu_i / d', \forall i \in [1, \ldots, C \cdot S]$.
```

**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.

**Algorithm 2** Rotation invariant $\nu$ description.

```
Require: a number of circles $C$, a number of sections $S$, and a set of bins $B$.
Ensure: the rotation invariant descriptor vector $\nu^k$.
  Define $G = \{G_1, \ldots, G_{S/2}\}$ as the $S/2$ diagonals of $B$, where $G_i = \{\nu(b_{\{1,i\}}), \ldots, \nu(b_{\{C,i\}}), \ldots, \nu(b_{\{1,i+S/2\}}), \ldots, \nu(b_{\{C,i+S/2\}})\}$
  Select $G_i$ so that $\sum_{j=1}^{2C} G_i(j) \geq \sum_{j=1}^{2C} G_k(j),\ \forall k \in [1, \ldots, S/2]$
  Define $L_G$ and $R_G$ as the left and right areas of the selected $G_i$ as follows:
  $L_G = \sum_{j,k} \nu(b_{\{j,k\}},\ j \in [1, \ldots, C],\ k \in [i+1, \ldots, i+S/2-1]$
  $R_G = \sum_{j,k} \nu(b_{\{j,k\}},\quad j \in [1, \ldots, C],\quad k \in [i+S/2+1, \ldots, i+S-1]$

  if $L_G > R_G$ then
```

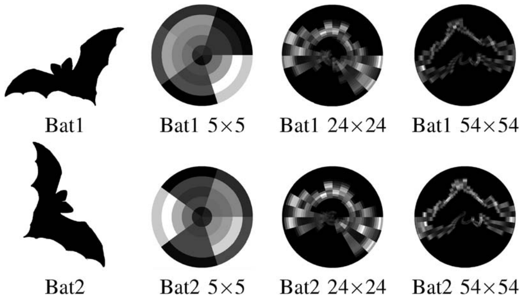


Fig. 3. Examples of image descriptors at different sizes for two object instances. The more regions used for the description, the more local information about the shape obtained. Notice that the two descriptors are correctly rotated and aligned.

```
    $B$ is rotated $k = i + S/2 - 1$ positions to the left
        $\nu^k = \{\nu(b_{\{1,k+1\}}),\ldots,\nu(b_{\{1,S\}}),\nu(b_{\{1,1\}}),\ldots,\nu(b_{\{1,k\}}),\ldots,$
        $\ldots,\nu(b_{\{C,k+1\}}),\ldots,\nu(b_{\{C,S\}}),\nu(b_{\{C,1\}}),\ldots,\nu(b_{\{C,k\}})\}$
else
    $B$ is rotated $k = i - 1$ positions to the right
        $\nu^k = \{\nu(b_{\{1,S\}}),\ldots,\nu(b_{\{1,S-k+1\}}),\nu(b_{\{1,1\}}),\ldots,\nu(b_{\{1,S-k\}}),\ldots,$
        $\ldots,\nu(b_{\{C,S\}}),\ldots,\nu(b_{\{C,S-k+1\}}),\nu(b_{\{C,1\}}),\ldots,\nu(b_{\{C,S-k\}})\}$
end if
```

A visual result of the rotation invariant process can be observed in Fig. 3 in which two bats with different descriptor orientations are rotated and aligned.

In this way, the output descriptor $\nu$ for an input region $I$ represents a distribution of the probabilities of the symbol structure considering the 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 Fig. 3, a bat instance from the public MPEG-7 data set [15] is described with different $C \times S$ correlogram sizes. Note that, when we increase the number of regions, the description becomes more local. Thus, the 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 the validation subset, defining the optimum number of sizes, encoding the different distortions on each particular problem, and offering the required tradeoff between the interclass and intraclass variabilities in a problem-dependent way.

The CBSM correlogram is defined by means of a number of sectors $S$ and a number of concentric circles $C$ in a linear correlogram design. It implies that the area of the outer sectors is higher than the area corresponding to the inner sectors. Since we define the same importance to all analyzed shape points, it seems intuitive to define the sectors with the same area. However, in this paper, we define a linear concentric circle definition which implies a more local description on the center of the 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 the symbols is usually higher compared to the inner variabilities (i.e., the external strokes in the hand-drawn symbols). On the other hand, if we want to define a correlogram structure where all the sectors have the same area, we simply need to change the distance among the correlogram sectors to satisfy the new constraints.

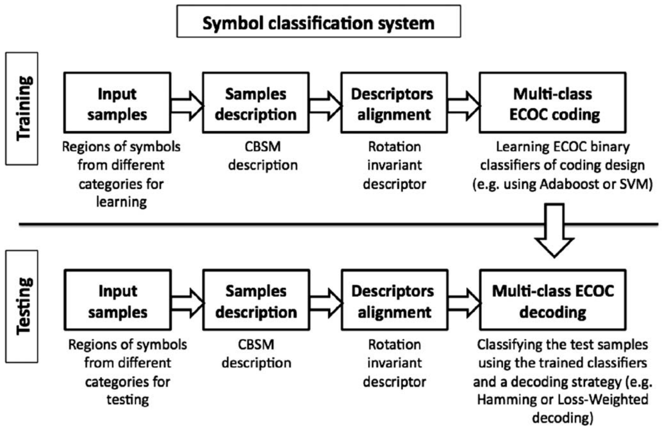

Symbol classification system
Training
Input samples
Samples description
Descriptors alignment
Multi-class ECOC coding
Regions of symbols from different categories for learning
CBSM description
Rotation invariant descriptor
Learning ECOC binary classifiers of coding design (e.g. using Adaboost or SVM)
Testing
Input samples
Samples description
Descriptors alignment
Multi-class ECOC decoding
Regions of symbols from different categories for testing
CBSM description
Rotation invariant descriptor
Classifying the test samples using the trained classifiers and a decoding strategy (e.g. Hamming or Loss-Weighted decoding)

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.

## 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 previously described rotation invariant CBSM descriptor is computed. For the second stage, the ECOC framework is used. The whole process is shown in Fig. 4.

ECOCs [13] are a metalearning strategy that divides a multiclass problem into a set of binary problems, solves them individually, and aggregates their responses into a final multiclass framework. The ECOCs have been successfully applied to many machine vision tasks [16]–[19], showing interesting properties in statistical learning, reducing both the bias and the variance of the base classifiers [20].

The ECOC metalearning 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 $\mathbf{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$, where each $\mathbf{x}_i$ belongs to a class $C_i \in \{C_1,\ldots,C_K\}$, the ECOC encoding consists of constructing $M$ binary problems using the original $K$ classes. Each binary problem splits into two metaclasses, and values $+1$ and $-1$ are assigned to each class belonging to the first and second metaclasses, respectively. If a class does not belong to any metaclass, 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 code word that is compared with each row of the matrix $\mathbf{T}$. The class code word with the minimum distance is selected as the classifier output. The ECOC scheme allows to represent in a common framework the well-known strategies, such as the one-versus-all or all-pair (one-versus-one) voting schemes, as well as the more sophisticated problem-dependent encodings, namely, the discriminant ECOC [21] or the subclass ECOC [14], without a significant increment of the code word length. The 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 code words. Escalera *et al.* [22] have recently shown that *weighted decoding* achieves the minimum error with respect to the most state-of-the-art decoding measures. The weighted decoding strategy decomposes the decoding step of the ECOC technique into two parts: a weighting factor for each code position and any general decoding strategy. In [22], Escalera *et al.* 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 code words. The complete decoding strategy weights the contribution of the decoding at each position of the code word by the elements of a weighting matrix $W$ that ensures that both conditions are fulfilled. As such, the final decoding strategy is defined as

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

where

$$w(i,j) = \begin{cases} r_i\left(S, \mathbf{T}(\cdot,j), \mathbf{T}(i,j)\right), & \mathbf{T}(i,j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$
$$\sum_{j=1}^{M} w(i,j) = 1,\ \forall i \in \{1,\ldots,K\}.$$

We define the metaclass relative accuracy ($r$ value) of class $k$ on the set $S$ given the definition of the metaclass $\rho$ as (1), shown at the bottom of the page, where $\rho$ defines which classes belong to which metaclass.

The second part of the weighting decoding relies on a base decoding strategy. In this paper, we use the linear loss-based decoding as the base decoding strategy. The linear loss-based decoding was introduced by Allwein *et al.* [23] and is defined as follows: Given the input sample $\mathbf{x}$ and the binary code $y$ which is the result of applying all the dichotomizers $(h_1, h_2, \ldots, h_M)$ to the input test sample, the decoding value is defined as

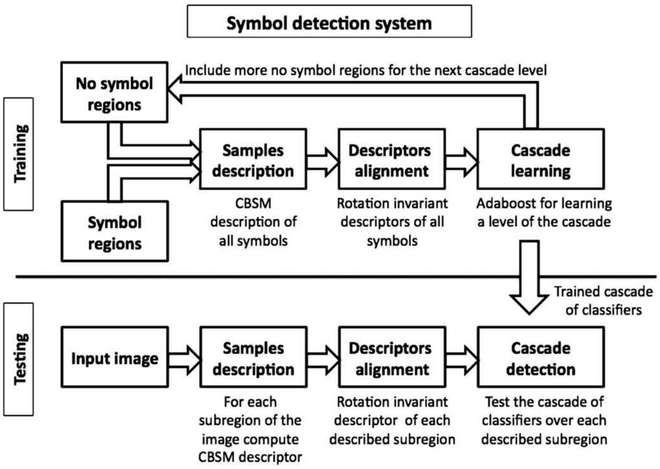

Symbol detection system
Training
No symbol regions
Include more no symbol regions for the next cascade level
Samples description
Descriptors alignment
Cascade learning
Symbol regions
CBSM description of all symbols
Rotation invariant descriptors of all symbols
Adaboost for learning a level of the cascade
Trained cascade of classifiers
Testing
Input image
Samples description
Descriptors alignment
Cascade detection
For each subregion of the image compute CBSM descriptor
Rotation invariant descriptor of each described subregion
Test the cascade of classifiers over each described subregion

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.

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

where $\mathbf{T}(i,\cdot)$ denotes the code word 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 misclassification. In the case of the linear loss-based decoding, we have $\mathcal{L}(\rho) = -\rho$.

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

### *B. Symbol Detection System*

In order to design a symbol detection methodology, two stages must be defined. The first stage (namely, the training) should learn to distinguish among the target object and *background* (i.e., learning a binary classifier). The second stage (namely, the 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 Fig. 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 and 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 the nonobject regions while

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

more complex analysis is performed in the 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 nonobjects are analyzed only until they are rejected by a stage. The number of applied classifiers is reduced exponentially due to the cascade architecture. This strategy is detailed in Algorithm 3.

**Algorithm 3** Attentional cascade training algorithm.

```
Require: a set of positive examples $P$, 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 classifiersh*.
  for$i = 1$ to $L$ do
    $F_i \leftarrow 1, n_i \leftarrow 0$
    while$F_i > f$ do
      $n_i \leftarrow n_i + 1$
      Use $P$ and $N$ to train a classifier with $n_i$ features
       using AdaBoost
      $F_i \leftarrow$ Evaluate the current cascaded classifier on
       the validation set
      Decrease the threshold for the $i$th classifier until
       the current cascaded classifier satisfies a detection
       rate of $a$ (this also affects $F_i$)
    end while
    $N \leftarrow 0$
    Evaluate the current cascaded detector on the set of
     nonobject images and put any false detections into
     the set $N$.
  end for
```

Once the cascade of classifiers is learned, 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: the target object regions $R$.
  for windows $W$ of size $S_I$, increasing by $i$ to $S_F$ do
    for each region $r$ in $I$ of size $W$ with shift $s$ among
     the regions, increasing by $i$ do
      test cascade $h$ over region $r$

$$h(r) = \begin{cases} 1 & \text{if detected as positive (object instance)} \\ & \text{save region} \rightarrow R = R \cup r \\ 0 & \text{if detected as negative (background)} \end{cases}$$

    end for
  end for
```

## IV. Experimental Evaluation

We divide the experimental evaluation into two main blocks: multiclass symbol categorization and symbol detection.

### A. Multiclass Symbol Categorization

In order to present the multiclass categorization results, we discuss the data, methods, and validation of the experiments.

1) *Data*: For comparing our CBSM multiclass methodology, we used two multiclass data sets: The first is the public 70-class MPEG-7[1] binary repository data set [15], which contains a high number of classes with different appearances of the symbols from the same class, including rotation. The second data set is a 17-class data set of grey-level symbols,[2] which contains the common distortions from real environments, such as the illumination changes, partial occlusions, or changes in the point of view.

2) *Methods*: The descriptors considered in the comparison results are the SIFT [26], BSM [10], Zoning [1], and Zernike moments [2]. The details of the descriptors used for the comparison results are discussed in the following sentences. The optimum correlogram size of the CBSM descriptor is estimated by applying a cross validation over the training set, using 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. The rotation invariance for the BSM descriptor is achieved by means of the principal component alignment before the descriptor computation [10]. Concerning the Zernike moment descriptor, seven moments are used. A 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 (LW) decoding [22] to solve the multiclass categorization problems. We also consider a support vector machine (SVM) with a radial basis function (RBF) base classifier for the ECOC design with $C = 1$ and $\gamma = 1$ and a three-nearest-neighbor (3-NN) classifier in the comparison results. The regularization parameter $C$ and the $\gamma$ 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 the replication of the experiments, although we are aware that this parameter might not be optimal for the analyzed data sets.

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

$$CI = \frac{1.96\, \sigma_{X_j}}{\sqrt{NT}} \tag{2}$$

[1] MPEG-7 Repository Database: http://www.cis.temple.edu/ latecki/research.html

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

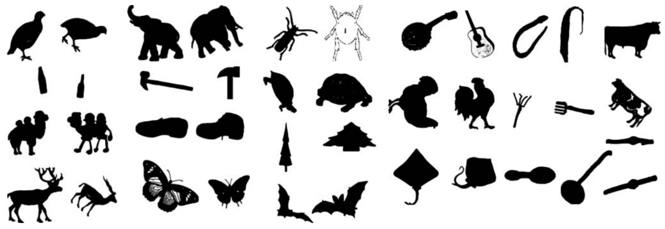


Fig. 6. MPEG-7 samples.

TABLE I
CLASSIFICATION ACCURACY AND CONFIDENCE INTERVAL (IN BRACKETS) ON THE 70 MPEG-7 SYMBOL CATEGORIES FOR THE DIFFERENT DESCRIPTORS USING A 3-NN CLASSIFIER AND THE ONE-VERSUS-ONE ECOC SCHEME WITH GENTLE ADABOOST AND RBF SVM AS THE BASE CLASSIFIERS

| Descriptor | 3 *NN* | 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) |

where $\sigma_{X_j}$ 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 the state-of-the-art descriptors over two multiclass categorization problems (with binary and grey-level symbols).

*1) MPEG-7 Multiclassification Data Set:* In this experiment, we used the 70 object categories from the public MPEG-7 binary object data set [15] to compare the whole set of descriptors in a multiclass categorization problem. A pair of samples of some classes of this data set are shown in Fig. 6.

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

*2) Grey-Scale Multiclassification 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. For this reason, we compare our CBSM with the SIFT descriptor [26] as well as with the BSM descriptor [10].

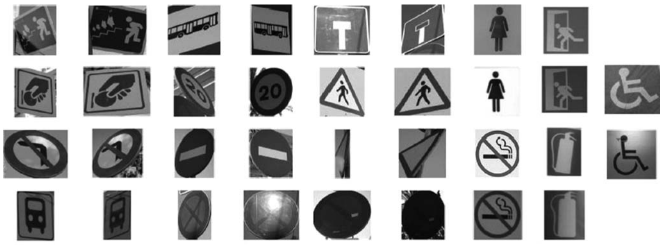


Fig. 7. Grey-scale symbol data set samples.

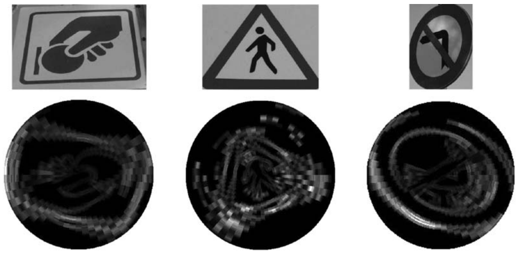


Fig. 8. CBSM descriptors from samples of the grey-level symbols data set.

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

| CBSM | BSM | SIFT |
|---|---|---|
| **77.82(6.45)** | 75.23(7.18) | 62.12(9.08) |

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 a one-versus-one ECOC scheme with Gentle AdaBoost as the base classifier and LW 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 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 the 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 the validation framework to measure the experimental evaluation.

1) *Data*: To test the detection CBSM methodology, we selected the predefined architectural plan files of the SmartDraw 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, Barcelona.[3]

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

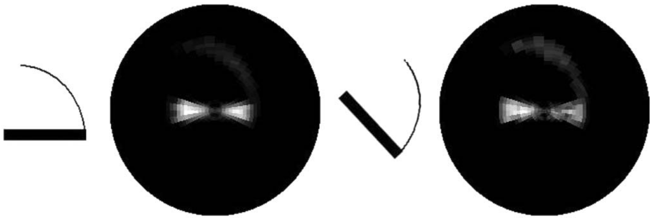


Fig. 9. Two examples of door positive images and their corresponding CBSM visual descriptors.

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

3) *Validation*: We apply the evaluation framework of Mikolajczyk *et al.* [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

$$1 - \frac{R_d \cap R_o}{R_d \cup R_o} < \epsilon \tag{3}$$

where $R_d$ is the detected region and $R_o$ is the original one. We set the maximum overlap error $\epsilon$ 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 with 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 the 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 SmartDraw 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 three positive detections with an intersection area greater than 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 the false positive isolated detection. Two examples of doors and their CBSM rotation invariant descriptors are shown in Fig. 9.

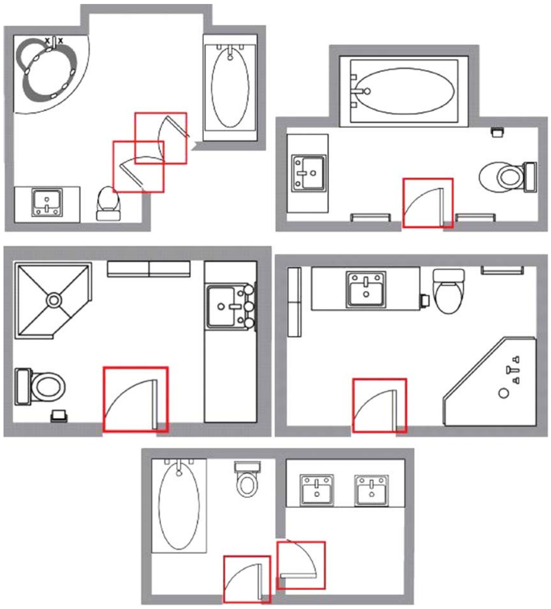
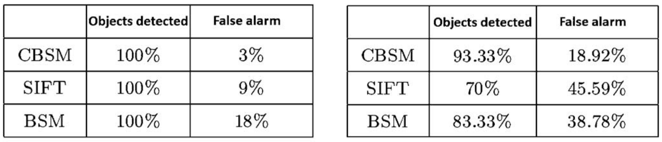


Fig. 10. SmartDraw architectural plan images and door symbol detection.



| | 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. 11. (a) Detection results over the architectural plan images. (b) Detection results over the musical score images.

Some visual results testing the CBSM detection procedure with a window shift of five pixels (which has been experimentally set) are shown in Fig. 10. 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 the three descriptors are shown in Fig. 11(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 (3), obtaining a hit ratio of 100%. Moreover, only one false positive region was detected in the case of the CBSM descriptor, corresponding to 3% of the detected regions. Note that one positive region from the thousands of analyzed regions is insignificant.[4]

*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 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 region as a positive region if there is a minimum of three intersections and discard the false positive isolated detection. Some results testing the CBSM detection procedure with a window shift of also five pixels on different staffs are shown in Fig. 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

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

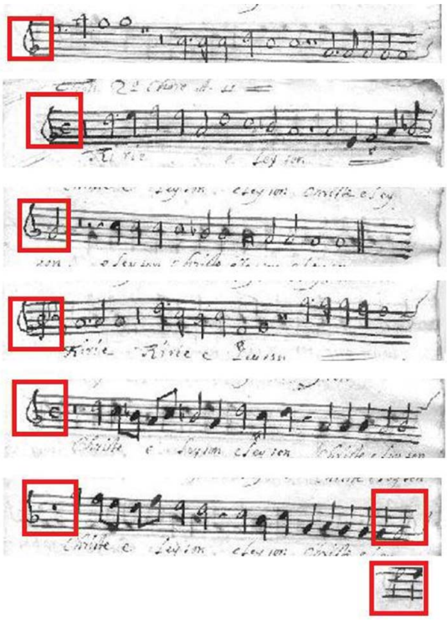


Fig. 12. Clef detection in old handwritten music score images. A false positive is shown at the bottom of the figure.

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 Fig. 11(b). In this case, the degradation of the images reduces the accuracy of the 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 (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 seven 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 CBSM descriptor has been presented. The new descriptor is suitable to describe and recognize, in a fast way, the symbols that can suffer from several distortions, such as occlusions, rigid or elastic deformations, discontinuities, or noise. The descriptor encodes the spatial arrangement of the 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 the object detection and multiclass categorization problems. In the case of multiclass symbol recognition, the new symbol descriptions are learned using the AdaBoost binary classifiers and embedded in an ECOC framework. The experimental results on different binary and grey-level multiclass categorization problems show that the CBSM descriptor obtains a higher performance than the state-of-the-art descriptors, particularly when classifying a high number of symbol classes that suffer from irregular deformations.

For the detection problem, the descriptor is learned using a cascade of classifiers with AdaBoost to discard the nonobject regions and tested over whole images, detecting the target objects. The symbol detection procedure presented in this paper has been shown to robustly locate the object instances in documents, such as the binary symbols in architectural plans and the 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 paper. 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 the different structure properties (such as those defined in [31]), and then, the CBSM descriptor could be defined from this new set of features.

## References

[1] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognit.*, vol. 37, no. 1, pp. 1–19, Jan. 2004.

[2] A. Khotanzad and Y. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990.

[3] W. Kim and Y. Kim, "A new region-based shape descriptor," Hanyang Univ./Konan Technol., Seoul, Korea, 1999.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[5] F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 1, pp. 34–43, Jan. 1986.

[6] B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. Hoboken, NJ: Wiley, 2002.

[7] H. Bunke, "Attributed programmed graph grammars and their application to schematic diagram interpretation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 6, pp. 574–582, Nov. 1982.

[8] J. Lladós, E. Martí, and J. J. Villanueva, "Symbol recognition by error-tolerant subgraph matching between region adjacency graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1137–1143, Oct. 2001.

[9] E. Valveny and E. Marti, "Hand-drawn symbol recognition in graphic documents using deformable template matching and a Bayesian framework," in *Proc. 15th Int. Conf. Pattern Recog.*, 2000, vol. 2, 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 Recognit. Lett.*, vol. 30, no. 15, pp. 1424–1433, Nov. 2009.

[11] K. Tombre, S. Tabbone, and P. Dosch, *Musings on Symbol Recognition*. New York: Springer-Verlag, 2006, pp. 23–34.

[12] D. Zuwala and S. Tabbone, *A Method for Symbol Spotting in Graphical Documents*. New York: Springer-Verlag, 2006, pp. 518–528.

[13] T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *J. Artif. Intell. Res.*, vol. 2, no. 1, pp. 263–286, Aug. 1995.

[14] S. Escalera, D. Tax, O. Pujol, P. Radeva, and R. Duin, "Subclass problem-dependent design of error-correcting output codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1041–1054, Jun. 2008.

[15] MPEG-7 Repository. [Online]. Available: http://knight.cis.temple.edu/~shape/MPEG7/dataset.html

[16] R. Ghani, "Combining labeled and unlabeled data for text classification with a large number of categories," in *Proc. IEEE Int. Conf. Data Mining*, 2001, pp. 597–598.

[17] T. Windeatt and G. Ardeshir, "Boosted ECOC ensembles for face recognition," in *Proc. Int. Conf. Visual Inf. Eng.*, 2003, pp. 165–168.

[18] J. Kittler, R. Ghaderi, T. Windeatt, and J. Matas, "Face verification using error correcting output codes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. 755–760.

[19] J. Zhou and C. Suen, "Unconstrained numeral pair recognition using enhanced error correcting output coding: A holistic approach," in *Proc. Int. Conf. Document Anal. Recognit.*, 2005, vol. 1, pp. 484–488.

[20] T. Dietterich and E. Kong, "Error-correcting output codes corrects bias and variance," in *Proc. Int. Conf. Mach. Learn.*, 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 Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 1001–1007, Jun. 2006.

[22] S. Escalera, O. Pujol, and P. Radeva, "On the decoding process in ternary error-correcting output codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 120–134, Jan. 2010.

[23] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Sep. 2002.

[24] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, vol. 28, no. 2, pp. 337–374, 2000.

[25] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2002.

[26] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[27] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New York: Wiley-Interscience, 2004.

[28] R. R. Johnson and P. Kuby, *Elementary Statistics*. Pacific Grove, CA: 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," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, Nov. 2005.

[31] V. Ferrari, F. Jurie, and C. Schmid, "Accurate object detection with deformable shape models learnt from images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–7.

**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 received the Ph.D. degree in multiclass visual categorization systems from the Computer Vision Center, UAB.

He is currently a Researcher with the Computer Vision Center. He is also currently a Lecturer with the Department of Applied Mathematics and Analysis, Universitat de Barcelona, Barcelona. His research interests include, between others, machine learning, statistical pattern recognition, visual object recognition, and human computer interaction systems.

**Alicia Fornés** received the B.S. degree from the Universitat de les Illes Balears, Palma, Illes Balears, Spain, in 2003, the M.S. degree from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 2005, and the Ph.D. degree in the writer identification of old music scores from the UAB in 2009.

She is currently a Postdoctoral Researcher with the Computer Vision Center, UAB. She is also currently with the Computer Science Department, UAB. 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 (UAB), Barcelona, Spain, in 2004.

He is currently an Associate Professor with the Department of Applied Mathematics and Analysis, Universitat de Barcelona, Barcelona. He is also currently with the Computer Vision Center, UAB. His research interests include the statistical machine learning techniques for object recognition and medical imaging analysis.

**Josep Lladós** received the degree in computer science from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 1991, and the Ph.D. degree from the Université Paris, Paris, France, in 1997.

Currently, he an Associate Professor with the Computer Science Department, UAB. He is also currently a Staff Researcher with the Computer Vision Center, UAB, where he is also the Director. His current research interests include document analysis, graphics recognition, and structural and syntactic pattern recognition.

Dr. Lladós is an active member of the International Association for Pattern Recognition (IAPR). He was the recipient of the IAPR-International Conference on Document Analysis and Recognition Young Investigator Award in 2007. He also created the company ICAR Vision Systems, a spin-off of the Computer Vision Center working on document image analysis, after being the recipient of the entrepreneurs award on business projects on Information Society Technologies from the Catalonia Government in 2000.

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

She is currently a Researcher with the Computer Vision Center, UAB. She is also currently an Associate Professor with the Department of Applied Mathematics and Analysis, Universitat de Barcelona, Barcelona. Her research interests include the development of physics-based and statistical approaches for object recognition, medical image analysis, and industrial vision.