Multi-class Binary Object Categorization using Blurred Shape Models

Sergio Escalera[1,2], Alicia Fornès[1,3], Oriol Pujol[1,2], Josep Lladós[1,3], and Petia Radeva[1,2]

[1]Computer Vision Center, Universitat Autònoma de Barcelona, Campus UAB, Edifici O, 08193, Bellaterra, Spain.
[2]Dept. Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007, Barcelona, Spain.
[3]Dept. Ciències de la Computació, Universitat Autònoma de Barcelona, Campus UAB, Edifici Q, 08193, Bellaterra, Spain.

Abstract. The main difficulty in the binary object classification field lays in dealing with a high variability of symbol appearance. Rotation, partial occlusions, elastic deformations, or intra-class and inter-class variabilities are just a few problems. In this paper, we introduce a novel object description for this type of symbols. The shape of the object is aligned based on principal components to make the recognition invariant to rotation and reflection. We propose the Blurred Shape Model (BSM) to describe the binary objects. This descriptor encodes the probability of appearance of the pixels that outline the object's shape. Besides, we present the use of this descriptor in a system to improve the BSM performance and deal with binary objects multi-classification problems. Adaboost is used to train the binary classifiers, learning the BSM features that better split object classes. Then, the different binary problems learned by the Adaboost are embedded in the Error Correcting Output Codes framework (ECOC) to deal with the muti-class case. The methodology is evaluated in a wide set of object classes from the MPEG07 repository. Different state-of-the-art descriptors are compared, showing the robustness and better performance of the proposed scheme when classifying objects with high variability of appearance.

Keywords. Shape descriptors, Multi-class classification, Adaboost, Error Correcting Output Codes.

1 Introduction

Shape recognition is one of the most popular areas of Pattern Recognition. Its aim consists in solving the problem of modeling and recognizing objects from a large set of classes. It is an extremely difficult task because of the high variability of the object appearance: changes in the perspective and viewpoint, occlusions, rigid and elastic deformations, and high intra-class an low inter-class variabilities. Several applications focus on this type of problems, such as the analysis of handwritten documents (e.g. analysis of old handwritten archive manuscripts

and sketching or calligraphic interfaces) [8]. A lot of effort has been made in the last decade to develop good symbol and shape recognition methods inspired in either structural or statistical pattern recognition approaches. In general, two major focus of interest can be stated: the definition of expressive and compact shape descriptors, and the formulation of robust classification methods according to such descriptors. Zhang [5] reviews the main techniques used in this field, mainly classified in contour-based descriptors (i.e. polygonal approximations, chain code, shape signature, and curvature scale space) and region-based descriptors (i.e. Zernique moments, ART, and Legendre moments [9]). A good shape descriptor should guarantee inter-class compactness and intra-class separability, even when describing noisy and distorted shapes. It has been shown that some object descriptors, robust to some affine transformations and occlusions in some type of objects, are not enough effective in front of elastic deformations. Thus, the research for other descriptors that address the problem of elastic and non-uniform distortions, as well as variations in object styles and blurring, is still required.

Concerning the categorization of objects' classes, many classification techniques have been researched based on both statistical or structural approaches. Elastic deformations of shapes modeled by probabilities tend to be learnt using statistical classifiers. One of the most well-known techniques in this domain is the Adaboost algorithm due to its ability for feature selection and its high performance when applied to binary problems [2]. Although many real problems require multi-classification, designing a single multi-class classifier remains a hard task. In such cases, the usual way to proceed is to reduce the complexity of the problem into a set of simpler binary classifiers and combine them in some way. An usual way to combine these simple classifiers is the voting scheme (one-versus-one or one-versus-all grouping schemes are the most frequently applied). In this way, Dietterich et. al. [11] proposed the Error Correcting Output Codes framework inspired in the signal processing coding and decoding techniques to benefit from error correction properties, obtaining successful results [10][11][3].

The goal of this paper is two-fold: on one hand, we introduce a novel shape descriptor, the Blurred Shape Model (BSM), that encodes the spatial probability of appearance of the shape pixels and their context information. As a result, a robust technique in front of elastic deformations is obtained. On the other hand, we present a successful scheme to describe and classify binary objects. The method aligns object shapes by means of the Hotelling transform and an area density adjustment. Then, the BSM is used for obtaining the shape description. The Adaboost algorithm is proposed to learn the descriptor features that best split classes, and the pairwise scheme (one-versus-one) with Error Correcting Output Codes increases the classification accuracy by correcting possible errors produced by the binary classifiers. A wide set of MPEG07 categories are described and classified with the present methodology, showing high success and better performance compared to the state-of-the-art descriptors.

The paper is organized as follows: Section 2 introduces the Blurred Shape Model descriptor. Section 3 presents the full binary object recognition scheme. Experimental results are shown in section 4, and section 5 concludes the paper.

2 Blurred Shape Model

The Blurred Shape Model (BSM) is based on the object shape description, allowing the definition of spatial regions where some parts of the shape can be involved.

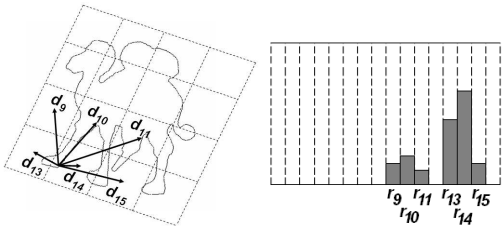
Given a binary image I,
Obtain the *shape* S contained in I
Divide I in $n \times n$ equal size sub-regions $R = \{r_i, ..., r_{n \times n}\}$, with c_i the center of coordinates for each region r_i.
Let $N(r_i)$ be the neighbor regions of region r_i, defined as $N(r_i) = \{r_k | r \in R, ||c_k - c_i||^2 \leq 2 \times g^2\}$, where g is the cell size.

For each point $\mathbf{x} \in S$,
For each $r_i \in N(r_{\mathbf{x}})$,
$d_i = d(\mathbf{x}, r_i) = ||\mathbf{x} - c_i||^2$
End_For
Update the probabilities vector v positions as:
$v(r_i) = v(r_i) + \frac{1/d_i}{D_i}, \quad D_i = \sum_{c_k \in N(r_i)} \frac{1}{||\mathbf{x} - c_k||^2}$
End_For
Normalize the vector v as: $v = \frac{v(i)}{\sum_{j=1}^{n^2} v(j)} \forall i \in [1, ..., n^2]$

Table 1. Blurred Shape Model algorithm.

Given a set of object shape points, they are treated as features to compute the BSM descriptor. The image region is divided in a grid of $n \times n$ equal-sized sub-regions (where the grid size identifies the blurring level allowed for the shapes). Each cell receives votes from the shape points in it and also from the shape points in the neighboring sub-regions. Thus, each shape point contributes to a density measure of its cell and its neighboring ones. This contribution is weighted according to the distance between the point and the center of coordinates c_i of the region r_i. Table 1 shows the algorithm.

In fig. 1, a shape description is shown for a MPEG07 sample. Figure 1(a) shows the distances of a shape point to the nearest sub-regions centers. To give the same importance to each shape point, all the distances to the neighbors centers are normalized. The output descriptor is a vector histogram v of length $n \times n$, where each position corresponds to the spatial distribution of shape points in the context of the sub-region and their neighbors ones. Fig. 1(b) shows the vector descriptor updating once the distances of the first point in fig. 1(a) are computed.



(a) (b)

Fig. 1. BSM density estimation example.

The resulting vector histogram, obtained by processing all shape points, is normalized in the range [0..1] to obtain the probability density function (pdf) of $n \times n$ bins. In this way, the output descriptor represents a distribution of probabilities of the object shape considering spatial distortions, where the distortion level is determined by the grid size. Referring the computational complexity, for a region of $n \times n$ pixels, the k relevant shape points considered to obtain the BSM require a cost of $O(k)$ simple operations.

3 Binary Object Recognition Scheme

In this chapter, we present the different methods applied in the scheme shown in fig. 2. First, we describe the Hotelling transform based on principal components and the area density readjustment for aligning the object shape. Then, we discuss the suitability of using Adaboost to train binary classifiers for the object classes and we comment the use of the Error Correcting Output Codes framework to extend the binary classification to the multi-class case.

Fig. 2. Process scheme.

3.1 Shape alignement

Before applying the proposed descriptor, a shape alignement process is performed. This process is composed of two steps: the first step, provides invariance to rotation by means of the Hotelling transform. And the second step deals with the posible mirroring effect.

The Hotelling transform finds a new coordinate system equivalent to locating the main axis of the object. Given a set of n representative object points defined as pairs of coordinates $\mathbf{x} = (x_i, y_i)$, where $i \in [1, .., n]$, the center of mass of

the object $m_{\mathbf{X}}$, and the eigenvectors V of the covariance matrix, the new transformation is obtained by means of the projection of the centered points of the object in the following way:

$$\mathbf{x}'_i = V(\mathbf{x}_i - m_{\mathbf{X}}), i \in [1, .., n] \quad (1)$$

Using this transform, we find the common axes for the different object instances. In fig. 3(a), the mean shape for the samples of one MPEG07 category after applying the Hotelling transform is shown. One can observe that the shapes are not properly aligned. For this reason, a second step, consisting of an area density estimation process is used. Observe fig. 3(b). Horizontal and vertical projections are applied to obtain the area of the object. Then, this area is projected on the two axes, as shown in fig. 3(b). The final alignment is obtained by horizontal and vertical reflection of the object in the direction of the higher area projections. The result of adjusting the alignment is shown in fig. 3(c). Another example of alignment for two MPEG07 object categories is shown in fig. 4.

(a) (b) (c)

Fig. 3. (a) Mean aligned shape based on principal components. (b) Horizontal and vertical area estimation. (c) Readjusted alignment.

Fig. 4. Mean aligned shapes for two MPEG07 categories.

3.2 Adaboost

Different types of objects may share local features [1] (see fig. 5). For this reason, Adaboost [2] has been chosen to boost the BSM models from different classes in order to define a classifier based on the features that best discriminate one class against another. Note that when comparing object descriptors, traditional matching distances take into account all object features for the final classification decision. When objects are very similar, slight deformations in the shared parts may include significant distance errors that finally can lead to a miss-classification of the objects. Observe fig. 5. The two objects have a discriminative region that splits the two categories (marked with a circle). Adaboost focuses on these regions by selecting the highest splitting features.

Fig. 5. Discriminant object regions.

3.3 Error Correcting Output Codes

The ECOC framework is a simple but powerful framework to deal with the multi-class categorization problem by embedding binary classifiers. Given a set of N_c classes, the basis of the ECOC framework consists of designing a codeword[1] for each of the classes. Arranging the codewords as rows of a matrix, a "coding matrix" M is defined, where $M \in \{-1, 0, 1\}^{N_c \times n}$, being n the code length. From the point of view of learning, M is constructed by considering n binary problems, each one corresponding to a matrix column. Joining classes in sets, each dichotomy defines a partition of classes (coded by +1, -1, according to their class set membership, or 0 if the class is not considered by the binary problem).

In figure 6(d), an example of a coding matrix M design is shown. The matrix is coded using 3 dichotomies $\{h_1, h_2, h_3\}$ for a three multi-class problem (c_1, c_2, and c_3). In fig. 6(a)-(c), three different sub-partition of classes are form, corresponding to all possible pairs of classes. This strategy is also called one-versus-one. Once we define the partitions of classes, each one is coded as a column of the coding matrix M, as shown in fig. 6(d). The dark regions are coded as +1 (first partition of classes), and the grey regions are coded as -1 (second partition of classes). The white regions correspond to the non-considered classes for their respective classifiers. Now, the rows of the matrix M define the codewords $\{Y_1, Y_2, Y_3\}$ for their corresponding classes $\{c_1, c_2, c_3\}$.

At the decoding step, applying the n trained binary classifiers, a code is obtained for each data point in the test set. This code is compared to the base codewords of each class defined in the matrix M, and the data point is assigned to the class with the "closest" codeword.

In fig. 6(e), an input test sample classification is shown. This input is tested using the three binary classifiers, and a codeword X is obtained. Finally, the Euclidean distance is applied between each class codeword and the test codeword X in the form $d(X, Y_i) = \sqrt{\sum_{j=1}^{n}(X(j) - Y_i(j))^2}$, where $i \in [1, .., 3]$. Finally, the test input X is classified by the class with minimum distance c_1.

An example of the process execution is shown in fig. 7. Fig. 7(a) shows an input image, which object shape is obtained in fig. 7(b) by means of a contour map. Shape alignment is performed by means of the Hotelling transform and the area density adjustment in fig. 7(d), and the final BSM of 32×32 grid size is shown in fig. 7(d).

[1] The codeword encodes the membership information of each binary problem for a given class.

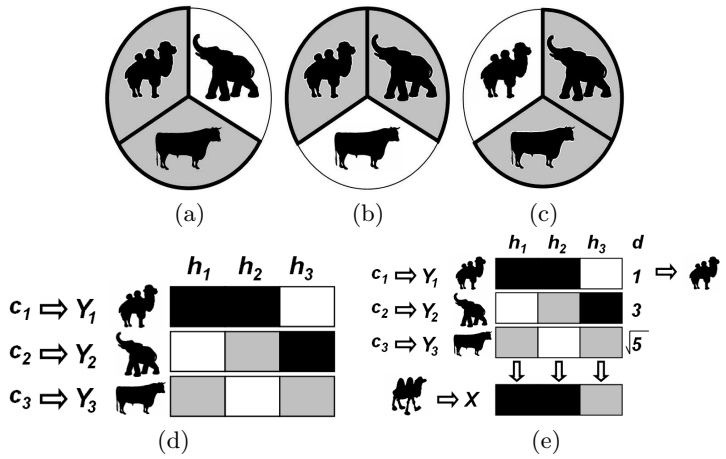


Fig. 6. (a)(b)(c) Three bi-partitions of classes for a three multi-class problem. (d) ECOC coding and (e) decoding for the problem.

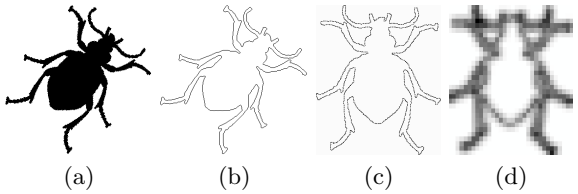


Fig. 7. (a) Input image, (b) contour map, (c) shape alignment, and (d) 32×32 BSM.

4 Results

To validate the system, first we describe the data, measurements, comparatives, and experiments.

Data: To test the system, we used 23 categories from the MPEG repository database [4]. This database has been chosen since it provides a high intra-class variability in terms of scale, rotation, rigid and elastic deformations, as well as a low inter-class variability. A pair of samples for each of the 23 categories are shown in fig. 8. Each of the classes contains 20 instances, which represents a total of 460 object samples.

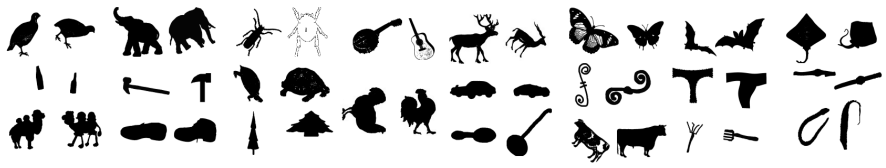


Fig. 8. MPEG07 classes.

Measurements: To analyze the performance of the techniques, the descriptors are trained using 50 runs of Discrete Adaboost with decision stumps, and the one-versus-one ECOC design with the Euclidean distance decoding. The classification score is computed by means of stratified ten-fold cross-validation with two-tailed t-test at 95% of the confidence interval.

Comparatives: The methods used for the comparative are: ART, Zoning, Zernique, and CSS curvature descriptors from the standard MPEG [7][5][6].

Experiments: To test the performance of the BSM descriptors, the comparative is applied over the set of 23 MPEG07 classes, classifying by means of 3-Nearest Neighbors to compare the descriptors robustness, and using the whole categorization system with Adaboost and ECOC to show its suitability for multi-class problems. Finally, we discuss the benefits of using the present methodology.

4.1 MPEG07 classification

The details of the descriptors used for the comparatives are the followings: BSM descriptor is of length 16×16 from the considered regions. The optimum grid size of 16×16 has been estimated applying cross-validation over the training set using a 10% of the samples to validate the different sizes of $n \in \{8, 12, 16, 20, 24, 28, 32\}$. The selected size is the one which attains the highest performance in the training set, defining the optimum grid encoding the blurring degree based on the database distortions. The scores obtained using cross-validation are shown in fig. 9. For a fair comparison, the Zoning descriptor is of the same size (16×16). The parameters for ART are radial order with value 2 and angular order with value 11. Concerning to Zernique, 7 moments are used to estimate the descriptor, and a length of 200 with an initial sigma of 1 increasing per one is applied for the curvature space of the CSS descriptor.

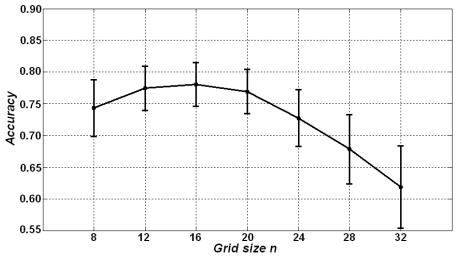
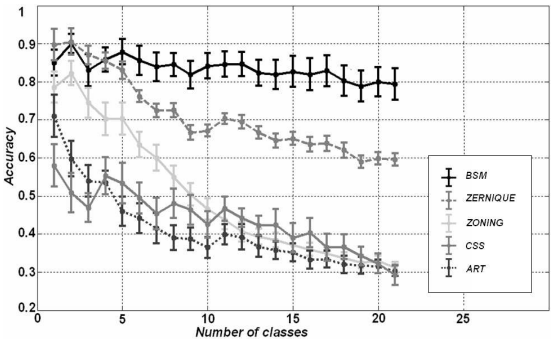


Fig. 9. Cross-validation on the training set for different BSM grid sizes.

For this experiment, we started the classification using the first 3 classes of fig. 8. Iteratively, one class is added at each step, and the classification is repeated until the 23 classes are processed. The main objective is to analyze the performance of the techniques when the number of classes increases. The results of the experiment are shown in fig. 10(a). Observing the figure, one can realize that the BSM descriptor attains the best performance for any number of classes in the classification system. Besides, an important point is that its performance

does not decrease significatively while increasing the number of classes, obtaining results around 80% in all cases. The second descriptor in the ranking is Zernique, which offers similar performance than BSM when the number of classes is small, but substantially decreases with the number of object categories. Finally, Zoning, CSS, and ART descriptors offers the worst classification scores in this problem. This can be intuitively justified by the fact that Zoning descriptors are very local, and the database is full of shape variations. This fact also affects to the CSS descriptor, since the points of curvature varies due to the high shape variations among objects.

In order to validate the descriptors independently of the system, the classification for the 23 MPEG07 classes is performed using a simple 3-Nearest Neighbors strategy based on the Euclidean distance. The results are shown in fig. 10(b). One can observe that for the different descriptors the performance decrease considerably. It is intuitively justified by the fact that all features contribute to the final decision, and the non-discriminative ones include distance errors that can miss-classify many samples. Nevertheless, observe that the reduction on the case of the BSM descriptor is less considerable, and it attains the best performance.



(a)

Descriptor	$3-NN$	Our system
BSM	**72.96±3.52**	**79.35±4.17**
Zernique	52.67±5.14	59.52±2.40
Zoning	21.74±2.91	32.07±1.81
CSS	29.50±5.65	28.85±2.42
ART	21.30±3.53	30.35±1.47

(b)

Fig. 10. (a) Classification accuracy on the MPEG07 object categories with our system. (b) Classification on the 23 MPEG07 object categories using 3-Nearest Neighbor and our system.

4.2 Discussion

Concerning the suitability of the presented scheme to deal with multi-class binary object categorization problems, several benefits should be mentioned:

The method is rotation invariant because of the use of the Hottelling transform and the area density adjustment. The method is also scaling and (x, y) stretching invariant because of the use of the $n \times n$ BSM grid. Besides, the BSM descriptor is robust against objects with rigid and elastic deformations since the size of the BSM grid defines the region of activity of the object shape points. The use of Adaboost as base classifier allows to learn difficult classes which may share several object features. The ECOC framework has the property of correcting possible classification errors produced by the binary classifiers, and allows

the system to deal with multi-class categorization problems. When the classifiers are trained only few features are selected, and when classifying a new test sample only these features are computed, which makes the approach very fast and suitable for real-time categorization problems.

5 Conclusions

We presented the Blurred Shape Model descriptor, which defines a probability density function of the shape of an object. The shape is parameterized with a set of probabilities that encode the spatial variability of the object, being robust to elastic deformations. Besides, a system to improve the performance of the novel descriptor dealing with multi-class categorization problems is proposed. Adaboost learns the discriminative features that better split object categories, and the binary classifiers are embedded in the Error Correcting Output Codes framework. The evaluation of the system is performed on 23 MPEG07 classes, showing great performance in cases of high intra-class and low inter-class variability, and outperforming the state-of-the-art descriptors while the computational cost is far less.

Acknowledgements

This work has been partially supported by the projects TIN2006-15694-C02-02 and TIN2006-15308-C02-01.

References

1. A. Torralba, K. Murphy, and W. Freeman, "Sharing visual features for multiclass and multiview object detection", Technical Report, Massachusetts Institute of Technology Computer Science and Artificial Intelligence (MIT AIM), 2004.
2. J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting", The Annals of Statistics, vol. 8, issue 2, pp. 337-374, 1998.
3. O. Pujol, P. Radeva, and J. Vitrià, "Discriminant ECOC: a heuristic method for application dependent design of error correcting output codes", in IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 28, pp. 1007-1012, 2006.
4. http://www.cis.temple.edu/ latecki/research.html
5. D. Zhang and G. Lu, "Review of shape representation and description techniques", Pattern Recognition, vol. 37, pp. 1-19, 2004.
6. W. Kim, "A new region-based shape descriptor", Technical report, Hanyang University and Konan Technology, 1999.
7. ISO/IEC 15938-5:2003(E)
8. J. Lladós, E. Valveny, G. Sánchez, and E. Martí, "Symbol Recognition: Current Advances and Perspectives", in Graphics Recognition: Algorithms and Applications, ed. D. Blostein and Y.B. Kwon, vol. 2390, Springer, Berlin, pp. 104-127, 2002.
9. B. Manjunath, P. Salembier, and T. Sikora, "Introduction to mpeg-7", Multimedia content description interface, John Wiley and Sons, 2002.
10. S. Escalera, O. Pujol, and P. Radeva, "Decoding of Ternary Error Correcting Output Codes", CIARP, Lecture notes in Computer Science, 2006.
11. T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes", Artificial Intelligence Research, vol. 2, pp. 263-286, 1995.