Handwritten Symbol Recognition by a Boosted Blurred Shape Model with Error Correction

Alicia Fornés, Sergio Escalera, Josep LLadós, Gemma Sánchez, Petia Radeva, and Oriol Pujol

Computer Vision Center, Dept. of Computer Science, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain

Abstract. One of the major difficulties of handwriting recognition is the variability among symbols because of the different writer styles. In this paper we introduce the boosting of blurred shape models with error correction, which is a robust approach for describing and recognizing handwritten symbols tolerant to this variability. A symbol is described by a probability density function of blurred shape model that encodes the probability of pixel densities of image regions. Then, to learn the most distinctive features among symbol classes, boosting techniques are used to maximize the separability among the blurred shape models. Finally, the set of binary boosting classifiers is embedded in the framework of Error Correcting Output Codes (ECOC). Our approach has been evaluated in two benchmarking scenarios consisting of handwritten symbols. Compared with state-of-the-art descriptors, our method shows higher tolerance to the irregular deformations induced by handwritten strokes.

1 Introduction

The analysis of handwritten documents has been a subject of intensive research for the last decades. The interest devoted to this field is not only explained from the scientific point of view, but also in terms of the social benefits that convey those systems. Two examples of interesting applications are the analysis of old handwritten archive manuscripts and sketching or calligraphic interfaces. The analysis of ancient documents is a growing interest in Europe and its main concern is not only the digitization but the extraction of knowledge from ancient documents to convert them to digital libraries, so that these documents can be edited and published, contributing to the diffusion and preservation of artistic and cultural heritage. Concerning to sketching interfaces, it is a joint interest between the fields of Pattern Recognition and Human Computer Interaction, which allows computers to integrate a natural way of interaction based on handwritten strokes which are interpreted as textual annotations or graphical gestures.

Although the analysis of textual handwritten documents has an intensive activity, the analysis of hand-drawn documents with graphical alphabets is an emerging subfield. Due to the fact that architectural, cartographic and musical documents use their own alphabets of symbols (corresponding to the domain-dependent graphic notations used in these documents), the automatic interpretation of such documents requires specific processes, within the field of Graphics Recognition, more than the field of Cursive Script Recognition. Two major differences between the two problems can be stated. Cursive script recognition has the context information in one dimensional way, but graphical alphabets usually are bidimensional. In addition, the use of syntactical knowledge, and lexicons, is more effective in text recognition than in diagrammatic notations because of the variability of structures and alphabets of the latter.

Symbol recognition is one of the central topics of Graphics Recognition [3]. A lot of effort has been made in the last decade to develop good symbol and shape recognition methods inspired in either structural or statistic pattern recognition approaches. The presence of handwritten symbols increases the difficulty of classification: there is a high variability in writing style, with different sizes, shapes and intensities, increasing the number of touching and broken symbols. In addition, working with old documents even increases the difficulties in these stages because of paper degradation and the frequent lack of a standard notation.

Symbol recognition in document images can be seen as a particular case of Shape Recognition. Two major focus of interest can be stated: the definition of expressive and compact shape description signatures, and the formulation of robust classification methods according to such descriptors. Zhang [7] 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. Zernike moments, ART, and Legendre moments [9]). A good shape descriptor should guarantee inter-class compacity and intra-class separability, even when describing noisy and distorted shapes. It has been proved that some descriptors, robust with some affine transformations and occlusions in printed symbols, are not efficient enough for handwritten symbols. Thus, the research of other descriptors for elastic and non-uniform distortions is required, coping with variations in writing style and blurring.

Concerning classification, numerous techniques (not necessary independent from each other) have been investigated based on statistical or structural approaches [3]. Elastic deformations of shapes modelled 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, detection, and classification problems [1]. Most classification algorithms are designed for multiclass problems. Nevertheless, this extension is normally hardly difficult. 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. 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). Dietterich et. al. [2] proposed a framework inspired in the signal processing coding and decoding techniques to benefit from error correction properties. The method is based on combining the weak classifiers as codified columns of a matrix and generate a codeword for each class. Thus, a test sample is evaluated with all the binary classifiers, and codewords are compared in the classification stage [2].

In this paper we present an approach to model and classify handwritten symbols. The method uses the context of the shape and defines a blurred region of the shape that makes the technique robust against elastic deformations (section 2). The Adaboost algorithm (section 3) is proposed to learn the descriptor features that best split classes, and the pairwise scheme (one-versus-one) with ECOC increases the classification accuracy by correcting possible weak classifiers errors. Finally, results (section 4) and the concluding remarks are exposed.

2 BSM: Blurred Shape Model

Handwritten symbol recognition is a hard task due to the high variability of symbol appearance because of the differences in writer sytles, and even also by the degradation in old documents (resolution, noise). The Blurred Shape Model (BSM) is based on the object shape parametrization, allowing the definition of spatial regions where some parts of the shape can be involved: Given a binary handwritten symbol, it is first skeletonized, and skeleton points are used as features to compute the BSM signature. The skeleton is applied to normalize the object shape in order to assign to each contour point the same importance and also to prevent different widths at some parts of the object. Then, the image is divided in a grid of $n \times n$ equal-sized subregions (where $n \times n$ identifies the blurring level allowed for the shapes). Each bin receives votes from the shape points in it and also from the shape points in the neighboring bins. Thus, each shape point contributes to a density measure of its bin and its neighboring ones. This contribution is weighted according to the distance between the point and the bin centroid of each neighbor.

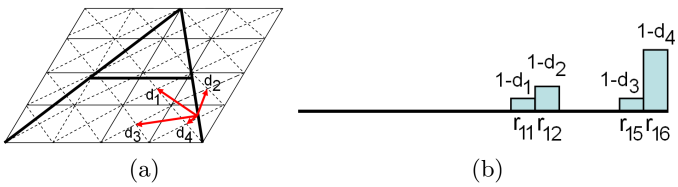


Fig. 1. (a) Shape pixel distances estimation respect to neighbor centroids. (b) Vector actualization of the region 16th, where d1+d2+d3+d4=1.

In Fig. 1, a letter shape parametrization is shown. Figure 1(a) shows the distances estimation of a shape point respect to the nearest centroids. To give the same importance to each shape point, all the distances to the neighbors centroids $\{d_1, d_2, d_3, d_4\}$ are normalized so that $d_1 + d_2 + d_3 + d_4 = 1$. The output descriptor is a vector histogram v of length $n \times n$, where each position corresponds to the amount of shape points in the context of the sub-region. The estimated normalized distances d_i for each affected sub-region r is used to actualize their corresponding vector locations adding the $1 - d_i$ values. Fig. 1 (b) shows the vector at this stage for the analyzed point of Fig. 1(a).

The resulting vector histogram, obtained by processing all feature 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. In Fig. 2, an input shape is processed. The symbol is filtered to obtain a thin shape, and the

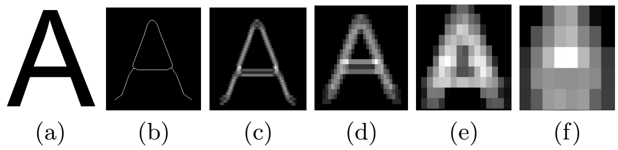


Fig. 2. (a) Input image. (b) Thinned image. (c) 64 regions blurred shape. (d) 32 regions blurred shape. (e) 16 regions blurred shape. (f) 8 regions blurred shape.

sequent figures correspond to the blurred parameterizations considering 64×64, 32×32, 16×16, and 8×8 sub-regions, respectively. The whole algorithm is summarized in Table 1.

Given a binary image I,
  Obtain the skeleton S of I
  Divide I in $n \times n$ equal size sub-regions

   for each point $(x, y) \in S$,
    let be $r_{x,y} \in R$ the sub-region containing (x, y),

     for $r_{x,y}$ and each $r'_{x,y} = \{r' \subset R | r'$ is neighbor of $r_{x,y}\}$
      $d_r = |cen(r), (x, y)|$

    Normalize each distance d_r as:
    $d_r = \frac{d_r}{\sum_{\forall i \in r'_i} d_i}$
    Actualize the probabilities vector v for $r_{x,y}$ and each $r'_{x,y}$ positions as:
    $v(r) = v(r) + (1 - d_r)$

  Obtain the blurred pdf normalizing 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. $|.|$ Is the Euclidean distance and $cen(r)$ is the centroid coordinates of the sub-region r.

3 Classification

In this section, the architecture of the classifier for the Blurred Shape Model descriptor and its benefits for handwritten symbols recognition is described. The whole process of the classification system is shown in Fig. 3.

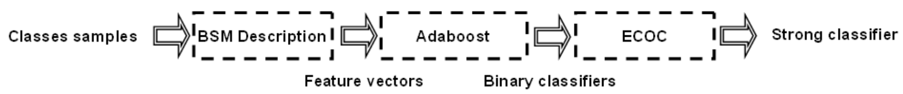
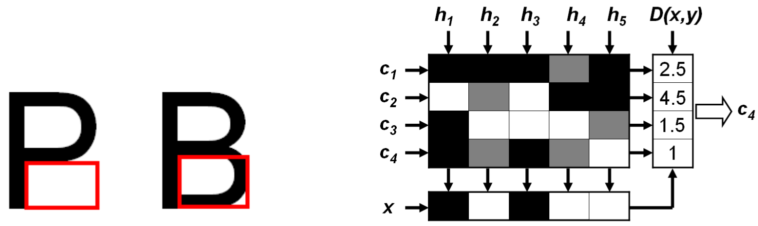


Fig. 3. Boosted blurred shape model with error correction scheme.

Adaboost [1] is used to train the classifier from Blurred Shape Model descriptors. The BSM has a probabilistic parametrization on the object shape considering its possible shape distortions. Different types of objects may share local features [4] (see Fig. 4(a)). For this reason, Adaboost has been chosen to boost the BSM model in order to define a classifier based on the features that best discriminate one classes against the others. In particular, we use the Discrete Adaboost version [1] with 50 iterations of decision stumps. To outperform the Adaboost behavior, we embed the Adaboost binary classifiers in the framework of Error Correcting Output Codes.



(a) (b)

Fig. 4. (a) Discriminate features for symbols that shares features. (b) Error Correcting Output Codes coding matrix for a 4 multiclass problem $\{c_1, ..., c_4\}$ using 5 binary classifiers $\{h_1, ..., h_4\}$. A test sample x is tested and classified by class c_4 applying the distance $D(x, y)$ between the test codeword and each class codeword.

The basis of the ECOC framework is to create a codeword for each of the N_c 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 (dichotomies), each 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 dichotomy). In Fig. 4(b) an example of a matrix M is shown. The matrix is coded using 5 dichotomies $\{h_1, ..., h_5\}$ for a four multiclass problem (c_1, c_2, c_3, and c_4). The white regions are coded by 1 (considered as positive for its respective dichotomy, h_i), the dark regions by -1 (considered as negative), and the grey regions correspond to the zero symbol (not considered classes for the current dichotomy). 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. 4(b), an input test sample x is shown. This input is tested using the five classifiers, and assigning the outputs to each codeword position (down of the figure). Finally, the hamming distance is applied between each class codeword and test codeword in the form $D(x, y) = \sum_{i=1}^{n} |x_i - y_i|/2$, where y is a class codeword, n is the number of classifiers, and $|.|$ is the absolute value. Finally the test input x is classified by the class at minimum distance c_4.

The ECOC framework shown increases the classification performance by the embedding of binary classifiers [2], [5]. In [8], it has been proved that the one-versus-one coding strategy outperforms the other traditional pre-defined coding strategies, and also that the Euclidean distance outperforms the traditional Hamming distance decoding. For this reason, the former strategy is used to evaluate the distance between the test sample and the class codewords. The estimation applied for each class candidate is $d(x,y) = \sqrt{\sum_{i=1}^{n}(x(i) - y(i))^2}$, where n is the code length, and x and y are the codewords for a test sample and a class.

4 Results

Two benchmarking databases of handwritten symbols have been used, namely clefs symbols from old musical scores, and architectural symbols from a sketching interface in a CAD framework. The database of clefs is obtained from a collection of modern and old musical scores (19th century) of the Archive of the Seminar of Barcelona. The database contains a total of 2128 samples between the three different types of clefs from 24 different authors. The images have been obtained from original image documents using a semi-supervised segmentation approach [6]. The main difficulty of this database is the lack of a clear class separability because of the variation of writer styles and the absence of a standard notation. In Fig. 5(a), one of the used old musical score is shown. The high variability of clefs' appearance from different authors can be observed in the segmented clefs of Fig. 5(b). The database of architectural hand-drawn symbols has 2762 total samples organized in the 14 classes shown in Fig. 6. Each class consists of an average of 200 samples drawn by 13 different authors. This database has been used to test the scalability of our method.

Fig. 5. (a) Old musical score, (b) High variability of clefs appearance: first row shows treble clefs, second row shows alto clefs and the third one shows bass clefs.

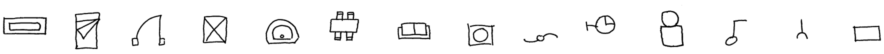


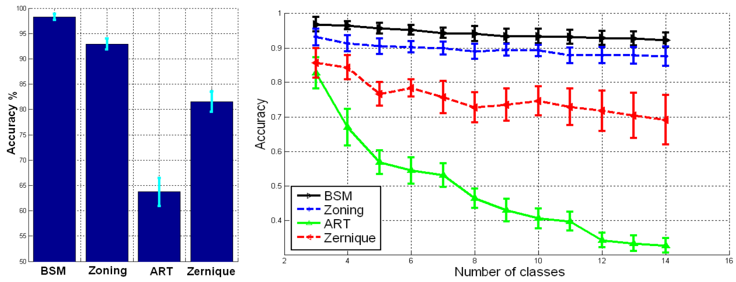
Fig. 6. Architectural handwriting classes.

To better assess the performance of our approach, it is compared with ART, Zoning, and Zernike descriptors [7][10]. The compared descriptors are also introduced in the classification framework to quantify the robustness of each descriptor at the same conditions. For all the experiments, stratified ten-fold cross-validation at 95% of the confidence interval is used. The descriptors for BSM and Zoning techniques are of length 8×8, from the considered sub-regions. The parameters for ART are radial order with value 2 and angular order with value 11; and for the Zernike descriptor, 7 Zernike moments are used.

The accuracy and confidence ranges results for the old musical score clefs are shown and graphically represented in Fig.7(a) and Fig.7(b), respectively. ART and Zernique descriptors obtain the minor results, while the Zoning descriptor in the classification scheme technique offers good results. The BSM strategy is the most robust, obtaining an accuracy upon 98%.

Method	Accuracy
BSM	98.11 ± 0.59
Zoning	92.45 ± 1.06
ART	63.75 ± 2.76
Zernike	81.50 ± 1.97

(a)



(b) (c)

Fig. 7. (a) and (b) Clefs classification results. (c) Descriptors classification accuracy increasing the number of architectural symbol classes.

The architectural symbol database has been used to test the performance under an increasing number of classes. We started the classification using the first 3 classes. Iteratively, one class was added at each step and the classification is repeated. The higher number of classes, the higher confusion degree among them because of the elastic deformations inherent to hand drawn strokes, and the higher number of objects to distinguish. The results of accuracy recognition in terms of an increasing number of classes are shown in Fig. 7(c). The performance of the ART and Zernike descriptors decreases dramatically when increasing the confusion in terms of the number of classes, while Zoning obtains higher performance. Finally, the accuracy of the BSM outperforms the other descriptors results, and its confidence interval only intersects with Zoning in few cases. This behavior is quite important since the accuracy of the latter descriptors remains stable, and BSM can distinguish the 14 classes with an accuracy upon 90%. Referring the computational complexity, for a region of $n \times n$ pixels, the $k \leq n \times n$ skeleton points are considered to obtain the BSM with a cost of $O(k)$ simple operations, which is faster than the moment estimation of the ART and Zernike descriptors. Besides, the Adaboost and ECOC strategies are very suitable for real-time multi-class classification problems [1].

5 Conclusions

We have presented the boosting of blurred shape models with error correction. A blurred shape model pdf is designed for each binary object, where the shape is parameterized with a set of probabilities that define the spatial invariance to elastic deformations of handwritten symbols. Adaboost learns the discriminative vector features, and the binary classifiers are embedded in the Error Correcting Output Codes framework. The evaluation of the technique in two real hand-written problems shows the outperforming of the novel methodology in comparison with the state-of-the-art descriptors and high robustness against elastic deformations. The skeleton information can also be changed to other structure criteria in the framework, allowing the context-based blurring of different object properties.

As future work, we are currently applying the symbol-based trained classifiers for symbol spotting in hand-written documents. Applying windowing techniques to image documents, regions can be described and evaluated by the classifiers in order to detect symbols. Besides, the detection can be done in real-time and speeded up by estimating only the features learned by Adaboost at each region.

Acknowledgements

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

References

1. 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.
2. T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes", *Artificial Intelligence Research*, vol. 2, pp. 263-286, 1995.
3. 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 of LNCS, Springer, Berlin, pp. 104-127, 2002.
4. 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.
5. S. Escalera, O. Pujol, and P. Radeva, "ECOC-ONE: A Novel Coding and Decoding Strategy", in *International Conference on Pattern Recognition* (ICPR), Hong Kong, vol. 3, pp. 578-581, 2006.
6. A. Fornés, J. Lladós, and G. Sánchez, "Primitive segmentation in old handwritten music scores", in *Graphics Recognition: Ten Years Review and Future Perspectives*, ed. W. Liu and J. Llados, vol. 3926 of LNCS, Springer-Verlag, pp. 279-290, 2006.
7. D. Zhang and G. Lu, "Review of shape representation and description techniques", *Pattern Recognition*, vol. 37, pp. 1-19, 2004.
8. O. Pujol and 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.
9. B. Manjunath, P. Salembier, and T. Sikora, "Introduction to mpeg-7", *Multimedia content description interface, John Wiley and Sons*, 2002.
10. W. Kim, "A new region-based shape descriptor", *Technical report, Hanyang University and Konan Technology*, 1999.