A New Adaptive Zoning Technique for Handwritten Digit Recognition

Sebastiano Impedovo, Francesco Maurizio Mangini, and Giuseppe Pirlo

Department of Computer Science, University of Bari "Aldo Moro"
Via Edoardo Orabona, 4 – 70125 Bari, Italy
francescomaurizio.mangini@uniba.it

Abstract. In this paper we present a new adaptive zoning technique based on Voronoi tessellation for the task of handwritten digit recognition. This technique extracts features according to an optimal zoning distribution, obtained by an evolutionary-strategy based search. Several experiments have been conducted on the MNIST and the USPS datasets to investigate the proposed approach. Comparisons with regular square zoning reveal that the presented zoning strategy achieves better results for any type of SVM classifier. Furthermore, the proposed zoning method shows that the combination of the adaptive zoning strategy with the Voronoi topology leads to find a distribution of zones able to improve accuracy significantly. As a matter of fact reached accuracies are close to the best algorithms.

Keywords: Zoning, Feature Extraction, Support Vector Machines.

1 Introduction

In handwritten character recognition, zoning is a well-known approach for feature extraction. It consists in dividing a pattern image into zones and collecting information from each zone. Whatever type of information is collected, the purpose of zoning is to allow to extract features from pattern images so that a given classifier can improve its accuracy. Many zoning methods have been proposed so far, based on different partitioning criteria [10].

Among the most recent approaches are those that attempt to adapt the distribution of zones depending on the patterns to be recognized. They have in common the idea to tailor zoning positions with the aim of meeting certain requirements. The most interesting works that fall into this category are briefly described below.

Radtke et al. [17] introduced an automatic approach to define zoning using Multi-Objective Evolutionary Algorithms (MOEAs). They provided a self adaptive methodology to find the best distribution of zones that obey to two optimality criteria: an error rate as low as possible and a minimal number of non-overlapping zones. They reported 95% accuracy with a nearest neighbor classifier on the NIST Special Database 19 (SD19).

Gatos et al. [8] adopted a zoning strategy that extracted features after adjusting the position of each zone according to local pattern information. For classification, they

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 91–100, 2013.

used the Euclidean distance between feature vectors combined with a minimum distance classifier. They achieved the best recognition accuracy of 88.35% on the CIL database of Greek handwritten characters, which consists of 28750 samples and has 46 classes.

Impedovo et al. [11] considered zoning design as a problem in which zone distribution had to be adapted in order to increase its discrimination capability. They reached an accuracy of 96% in handwritten digit recognition on the CEDAR dataset.

In this paper, inspired by the cited works, a new adaptive zoning technique based on Voronoi tessellation is presented. This approach tackles zoning design as an optimization problem and the optimal zoning distribution is obtained by an evolutionary strategy [1]. It differs from [11] in the problem definition, the optimization strategy, as well as the classifier.

Since SVM is a state-of-the-art classification method, accuracy performances of the proposed zoning approach have been experimentally investigated with several standard SVM classifiers. In addition, these performances have been compared with the ones obtained by regular square zoning.

Experimental tests have been conducted on the MNIST and the USPS databases of handwritten numerals. The results show that the proposed zoning method outperforms regular square zoning for any type of SVM classifier.

The remaining part of this paper is organized as follows. Section 2 formulates the zoning problem. Section 3 describes the proposed technique for zoning design. The experimental tests and the results are discussed in Section 4. The conclusions are drawn in Section 5.

2 Zoning Problem Definition

The introduced zoning strategy is based on Voronoi tessellation. It is a set of zones, called Voronoi polygons, built with respect to M generator points enclosed in the plane. Each generator p_i lies inside the boundaries of a Voronoi polygon Pol_i with the following property:

$$Pol_i = Pol(p_i) = \{p | d(p_i, p) \leq d(p_j, p), i \neq j\} \quad (1)$$

where $d(p_i, p_j)$ is the distance from point p_i to p_j.

Pol_i is the set of points closer to p_i than to any other p_j. The tessellation is $D_{Voronoi} = \{Pol_i\}_{i=1,..M}$ (M being the number of zones).

For a given set of generator points, $D_{Voronoi}$ is unique and generates regions which are path connected. Hence, the Voronoi Tessellation can be a zoning method that partitions the image into a set of Voronoi polygons, given a set of distinct points into the image plane. In particular, if I is an image of WxH pixels, a Voronoi diagram refers to a tessellation of the domain $I = \{(x, y) | 0 \leq x \leq W, 0 \leq y \leq H\}$ by the Voronoi regions $\{Pol_i\}_{i=1,..M}$ associated with a set of given generators $\{p_i\}_{i=1,..M}$. Here, the Euclidean distance has been used in (1).

The aim of our approach is to find the image partition that improves the accuracy of a given classifier. This can be viewed as an optimization problem $\mathcal{P}(S, \mathcal{F})$, where S is the search space, whose elements are candidate solutions, and $\mathcal{F}: S \rightarrow R$ is the function to be minimized. In our case, S consists of all the possible M-permutations without repetition of all the n pixel points $p_i = (x_i, y_i)$ belonging to the image I, being M the number of zones and n the number of pixels in the image. According to our purpose, the objective function $\mathcal{F}$ is assumed equal to the misclassification error rate *Err* committed by a given classifier:

$$\mathcal{F}(S) = Err(S) \tag{2}$$

If S is the set of all the Voronoi generating points $\{p_i\}_{i=1,..M}$, with $p_i \in I$, the minimization problem $\mathcal{P}(S, \mathcal{F})$ can be written as follows:

Find $s^* \in S$, $s^* = \{p^*_i\}_{i=1,..,M}$ so that:

$$\mathcal{F}(s^*) = \min_{s \in S} \mathcal{F}(s) \tag{3}$$

where p_i are the object variables of the optimization problem.

In this paper, for the optimization problem (3), an evolutionary approach has been used [6] because evolution strategies are generally known to be quite fast and efficient solvers.

3 The Evolutionary Strategy

Evolutionary strategies [1, 6] are iterative optimization techniques which try to discover an optimal solution through stochastic small changes of the object variables.

Figure 1 depicts the evolution cycle of the presented Evolutionary Strategy as a flowchart. This flowchart starts from an initial candidate solution s^0 that enters an evolution loop. The parent is then used for generating a set of λ offsprings by the mutation operator (4).

Thereafter, the population of the next generation is selected as the μ individuals with the best fitness, either from the offsprings and the parent (plus strategy μ+λ). This loop is repeated until a termination criterion is met.

If p_i^g and p_i^{g+1} are respectively the point p_i at g-th and (g+1)-th generation, the mutation involves replacing each point $p_i^g = (x_i^g, y_i^g)$ with $p_i^{g+1} = (x_i^{g+1}, y_i^{g+1})$ as follows:

$$\begin{cases} x_i^{g+1} = x_i^g + \Delta_{x_i^g}(0, \sigma^g) \\ y_i^{g+1} = y_i^g + \Delta_{y_i^g}(0, \sigma^g) \end{cases} \quad \forall\, i = 1,..,M \tag{4}$$

where:

- $\Delta_{x_i^g}(0, \sigma^g)$ and $\Delta_{y_i^g}(0, \sigma^g)$ are random Gaussian numbers with a mean of zero and standard deviation of σ^g.
- σ^g is the standard deviation at g-th generation.

The standard deviation plays a key role as it allows to control the speed of convergence. It is often called step size or mutation strength and its adjustment is one of the most important factors in the algorithm. In this work, the control of standard deviation is performed by log-normal self-adaptation operator [1]:

$$\sigma^{g} = \sigma\, e^{\tau\, N(0,\,1)} \tag{5}$$

Equation (5) contains a new fundamental parameter to be fixed: the learning rate τ. The general recommendation is to choose $\tau \propto 1/\sqrt{M}$, as it has been proven to be optimal with respect to the convergence speed in a theoretical case [6].

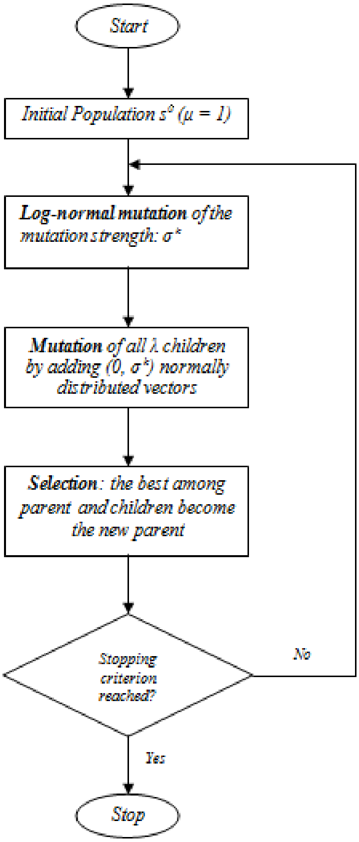


Fig. 1. Flow chart of the evolutionary approach

There is not an exact rule to decide the size of the parent population. Several theoretical studies [1] have been conducted and reported in the literature on the estimation of the optimal ratio λ/μ for $(1, \lambda)$ strategies. They showed that, although the optimal ratio depends on the objective function and its complexity, a ratio λ/μ equal to 5 can be a good starting point.

The convergence criterion is based on the evolution of the best value of the objective function along generations. The optimum is assumed to be reached as far as the best value has not significantly changed in the last generations. If F_{min}^{g} and F_{min}^{g+1} indicate the minimal values of the objective function inside the g-th and (g+1)-th generation, the iterative process will be stopped if:

$$| F_{min}^{g} - F_{min}^{g+1} | < \varepsilon \tag{6}$$

Moreover, a maximum number of iterations N_{max} is defined, after which the iterative process is stopped even though the stopping rule (6) has not been satisfied.

4 Experimental Results

Experiments have been carried out using regular square zoning and the presented zoning technique (called "Voronoi-based zoning" in the following), combined with a feature set composed of gradient-based and pixel density features.

The MNIST and the USPS datasets have been used. The former contains 60000 images of digits 0 − 9 for training and 10000 images for testing. The latter contains 7291 images of digits 0 − 9 for training and 2007 images for testing. The MNIST images are 28 × 28 grayscale images, while the USPS images are 16 × 16 pixels sized grayscale images.

On a trial and error basis, the number of zones has been chosen equal to 25. A preprocessing phase has been performed to generate the input digit images. It has been consisted of binarization using ImageJ software [19], and size normalization into 30×30 images for the MNIST dataset and 20×20 images for the USPS dataset.

According to the mentioned schemes, gradient features complemented by information on pixel density have been extracted from each zone.

Gradient features consist of calculating gradient values and constructing local histograms along certain orientations. As each pixel of the image has an orientation and magnitude depending on the local gradient, histograms have been built by accumulating magnitude values over the same directions for all the pixels belonging to each zone of the image. The gradient computation has been performed by the Sobel operator. The input image has been convolved with Sobel masks to calculate the gradient components in the horizontal $g_x(p)$ and vertical direction $g_y(p)$ at each pixel p. The magnitude G(p) and the angle Φ(p) of the pixel p have been given by :

$$G(p) = \sqrt{g_x^2(p) + g_y^2(p)} \tag{7}$$

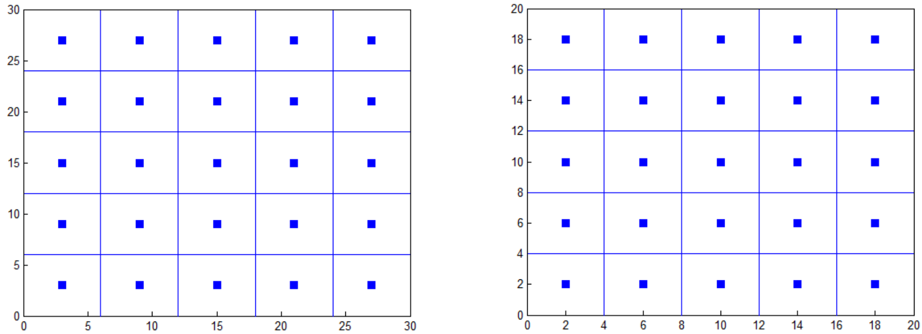
$$\Phi(p) = \mathrm{atan2}\,(g_x(p), g_y(p)) \in [0, 2\pi[\tag{8}$$

The gradient direction at each pixel has been decomposed into components according to the 8-directional chain code.

Pixel density has been computed as the number of black pixels that stand in each zone over the total number of pixels in the image. These complementary features have been added to the gradient ones, because they were able to further improve the accuracy. Therefore, we obtained a feature vector of 225 elements for each handwritten numeral.

All the experiments have been done in MATLAB version 7.8 (Release 2009a) 64 bits. It has been used an Intel Core i5 (64 bits), running on a Windows 7 Professional 64 bits Operating System, with 8 GB of RAM memory and 2 TB of disk space.

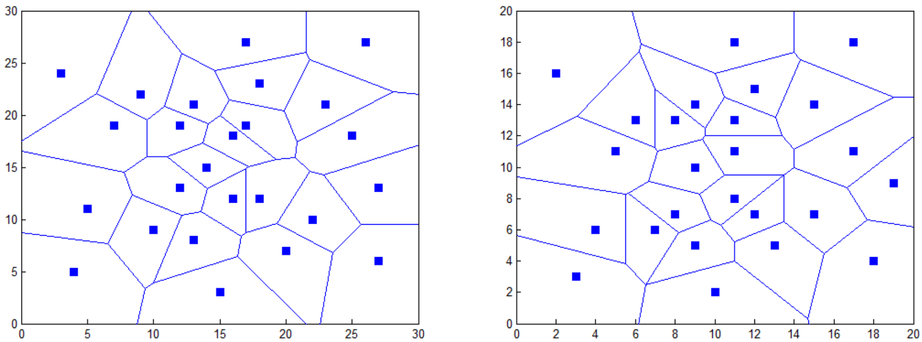
The first experiment has been dedicated to find the optimal zoning distribution for both the MNIST and the USPS datasets. To this purpose, the optimization problem (3) has been tackled adopting the *plus* strategy with $\mu = 1$. The initial parent population s^0 has been chosen as the set of Voronoi generating points $\{p^0_i\}_{i=1,..,M}$ to which corresponds the regular square zoning (see Figure 2).



(a) (b)

Fig. 2. Initial population s^0 for the MNIST (a) and the USPS (b) datasets

This choice has been taken because no assumptions can be done about the optimal distribution of zones. The objective function $\mathcal{F}(s)$ (2) has been assumed as the k-fold cross validation error (k=5), computed by a linear SVM classifier. For this classifier, the value of the hyperparameter C has been set to 10.



(a) (b)

Fig. 3. Optimal zoning distribution for (a) the MNIST and (b) the USPS datasets

The following working parameters have been selected for the evolution strategy: $\lambda = 5$, $\sigma = 1$, $\tau = 0.8$, $N_{max} = 40$, $\varepsilon = 0.01$.

Figure 3 shows that the MNIST and the USPS optimal zoning configurations are quite similar. As a matter of fact, it can be observed that for both cases there is a higher concentration of Voronoi generator points in the center part of the image, while the edge points are less dense and have rather similar positions.

These zoning distributions are the ones which allow to separate in the most effective way patterns belonging to different classes, while bringing together those belonging to the same class.

From the previous figure, it comes out that there is a set of generating points able to discriminate effectively handwritten numeral shapes, according to the proposed zoning method and the selected features.

Intuitively, their fine positions depend on people's handwriting styles and most probably this is the reason why the two sets are not completely overlapped.

In order to evaluate the performances of the Voronoi-based zoning, several standard SVM classifiers were trained on the features extracted by these zoning distributions:

$$\Phi_{\mathrm{LIN}}(\mathbf{x_i}, \mathbf{x_j}) = \mathbf{x_i} \cdot \mathbf{x_j} \tag{9}$$

$$\Phi_{\mathrm{POLY}}(\mathbf{x_i}, \mathbf{x_j}) = (1 + \mathbf{x_i}^T \cdot \mathbf{x_j})^5 \tag{10}$$

$$\Phi_{\mathrm{RBF}}(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma \| \mathbf{x_i} - \mathbf{x_j} \|^2) \tag{11}$$

They are, respectively, a linear, a degree 5 polynomial and a radial basis function kernel SVM. It has been chosen the polynomial degree d = 5, because it has been proven to have the best performance [14].

Using LIBSVM [3], one-against-one SVM classifiers have been trained for each possible pair of classes. Then, a test pattern has been assigned by evaluating all binary classifiers and performing voting among them [20].

The values of C and γ used to perform classification are summarized in the following table (Table 1).

Table 1. Parameters used to perform classification

	MNIST	**USPS**
C	10	10
γ	0.125	0.004

The combination of C and γ has been selected assuming C = 10 (default value) and calculating γ on a trial-and-error basis. So, it must be underlined that we did not perform any parameter optimization for these experiments.

Table 2 reveals that Voronoi-based zoning always outperforms regular square zoning for any type of SVM classifier.

In particular, even the performance of the linear kernel is remarkably improved for both datasets. It means that Voronoi-based zoning, combined with the feature we chose, is able to enhance the performance of SVM classifiers.

Besides, it can be observed that RBF and polynomial kernel SVM accuracy are competitive with state-of-the-art approaches.

Table 2. Accuracy on the test sets (%)

	MNIST		USPS	
	Regular Square Zoning	**Voronoi - based Zoning**	**Regular Square Zoning**	**Voronoi - based Zoning**
$\mathbf{SVM}_{LIN}$	96.71	97.86	95.07	96.46
$\mathbf{SVM}_{POLY}$	98.19	99.18	95.81	96.96
$\mathbf{SVM}_{\mathbf{RBF}}$	98.26	99.23	96.21	97.01

For the MNIST dataset, the best result available in the literature is 99.81% using a Convolutional Neural Network as a trainable feature extractor and a RBF kernel SVM as a recognizer [15]. Therefore, our results are not too far from [15] that achieves a better accuracy value, but uses a more sophisticated feature extraction algorithm.

For the USPS dataset, the human recognition rate estimated to be 97.5% [2] highlights that it is a quite hard database. In this case, some of the best results available in the literature are those obtained by Simard et al. [5, 22]. They reached 97.4% accuracy on the test set using a modified training set, that had been enhanced by adding machine-printed characters. Hence, once again, our results are noteworthy, since we employed the original training set.

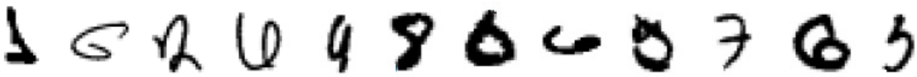


Fig. 4. Examples of some MNIST misclassified digits

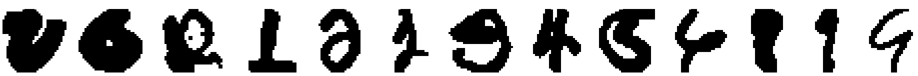


Fig. 5. Examples of some USPS misclassified digits

Anyway, our recognition results on the MNIST and the USPS test sets are a little bit lower than the best results reported in the literature. Notwithstanding this, our goal in this paper was not to reach the maximum recognition rate, but instead to find a zoning distribution capable of improving performance significantly. Thus, it can be concluded that the combination of the adaptive approach with the Voronoi topology leads to achieve excellent results. The adaptive approach can discover the distribution of zones that maximizes the interclass differences while minimizing the intraclass ones. Furthermore the Voronoi topology, allowing variations in the shape of the zones, introduces an additional degree of freedom useful to the same purpose.

The classification times of our method are shown in Table 3. They are measured in seconds.

Table 3. Classification times (s)

	MNIST	USPS
SVM_{LIN}	19	5
SVM_{POLY}	61	13
SVM_{RBF}	223	46

Hence, it can be observed that the polynomial SVM classifier with polynomial degree d = 5 represents a good trade-off in terms of both accuracy and classification speed.

5 Conclusions

In this paper, we presented a new zoning technique to address the handwritten digit recognition problem. Our technique considered zoning design as an optimization problem and used an evolutionary strategy to find the optimal zoning distribution. Our method is able to adaptively determine the optimal position of the Voronoi generating points.

The comparison of zoning method performance for handwritten numeral recognition is a difficult task since there are differences in experimental methodology and settings, as well as differences in the databases used. Here, in order to test the proposed approach, several standard SVM classifiers were trained on the features extracted by this zoning distribution. We chose SVM classifiers as they have been successfully used in character recognition.

The experiments show that this zoning method achieves a general improvement in performance, for any type of SVM classifier. Hence, this work demonstrates the utility of adaptive zoning and Voronoi topology for SVM classification of handwritten digits. In particular, our zoning strategy led to a recognition rate of 99.23% on the MNIST dataset and of 97.01% on the USPS dataset.

Future works include investigation on the effectiveness of our method by using larger features, in order to improve recognition rate without a significant increase of the classification time.

References

1. Beyer, H.-G.: The Theory of Evolution Strategy. Springer, New York (2001)
2. Bromley, J., Sackinger, E.: Neural-network and k-nearest-neighbor-classifiers. TechnicalReport 11359-910819-16TM, ATT (1991)
3. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm
4. de Berg, M., Cheong, O., van Kreveld, M., Overmars, M.: Computational Geometry: Algorithms and Applications. Springer, Berlin (2008)
5. Drucker, H., Schapire, R., Simard, P.: Boosting performance in neural networks. International Journalof Pattern Recognition and Artificial Intelligence 7, 705–719 (1993)

6. Fogel, D.B.: An introduction to simulated evolutionary optimization. IEEE Transactions on Neural Networks 5(1), 3–14 (1994)
7. Freitas, C.O.A., Oliveira, L.S., Bortolozzi, F.: Handwritten Character Recognition using nonsymmetrical perceptual zoning. IJPRAI 21(1), 1–21 (2007)
8. Gatos, B., Kesidis, A., Papandreou, A.: Adaptive Zoning Features for Character and Word Recognition. In: Proceedings of ICDAR 2011, pp. 1160–1164 (2011)
9. Hull, J.: A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 550–554 (1994)
10. Impedovo, D., Pirlo, G., Modugno, R.: New Advancements in Zoning-Based Recognition of Handwritten Characters. In: Proceedings of ICFHR 2012, pp. 661–665 (2012)
11. Impedovo, S., Modugno, R., Pirlo, G.: Analysis of membership Functions for Voronoi-based Classification. In: Proceedings of ICFHR 2010, pp. 220–225 (2010)
12. Impedovo, S., Modugno, R., Pirlo, G.: Membership Functions for Zoning-based Recognition of Handwritten Digits. In: Proc. International Conference on Pattern Recognition, Istanbul, Turkey, pp. 1879–1879 (2010)
13. LeCun, Y., et al.: Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, vol. 86, pp. 2278–2324 (1998)
14. Liu, C.-L., Nakashima, K., Sako, H., Fujisawa, H.: Handwritten digit recognition: benchmarking of state-of-the-art techniques. Pattern Recognition 36, 2271–2285 (2003)
15. Niu, X.X., Suen, C.Y.: A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recognition 45, 1318–1325 (2012)
16. Oliveira, L.S., Sabourin, R., Bartolozzi, F., Suen, C.Y.: Automatic Recognition of Handwritten Numeral Strings: A Recognition and Verification Strategy. IEEE T-PAMI 24(11), 1438–1454 (2002)
17. Radtke, P.V.W., Wong, T., Sabourin, R.: A Multi-objective Memetic Algorithm for Intelligent Feature Extraction. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) EMO 2005. LNCS, vol. 3410, pp. 767–781. Springer, Heidelberg (2005)
18. Rajashekararadhya, S.V., Ranjan, P.V.: Support Vector Machine based Handwritten Numeral Recognition of Kannada Script. In: Proc. IEEE International Conference on Advance Computing 2009, pp. 381–386 (2009)
19. Rasband, W.S.: ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/
20. Schölkopf, B., Smola, A.: Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge (2002)
21. Sharma, D., Gupta, D.: Isolated Handwritten Digit Recognition using Adaptive Unsupervised Incremental Learning Technique. International Journal of Computer Applications 7(4), 27–33 (2010)
22. Simard, P., LeCun, Y., Denker, J.S.: Efficient pattern recognition using a new transformation distance. In: Advances in Neural Information Processing Systems 5, San Francisco, CA, USA, pp. 50–58 (1993)
23. Vamvakas, G., Gatos, B., Pratikakis, I., Stamatopoulos, N., Roniotis, A., Perantonis, S.J.: Hybrid off-line OCR for isolated handwritten Greek characters. In: Proc. 4th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2007), Innsbruck, Austria, pp. 197–202 (2007)
24. Xiang, P., Xiuzi, Y., Sanyuan, Z.: A hybrid method for robust car plate character recognition. Engineering Applications of Artificial Intelligence 18(8), 963–972 (2005)