A Combined Statistical-Structural Strategy for Alphanumeric Recognition

N. Thome and A. Vacavant

LIRIS - UMR 5205, Université Lumière Lyon 2
5, avenue Pierre Mendès-France
69676 Bron cedex, France
{nicolas.thome,antoine.vacavant}@liris.cnrs.fr

Abstract. We propose an approach dedicated to recognize characters from binary images by an hybrid strategy. A statistical method is used to identify the global shape of each alphanumeric symbol. The recognition is managed by a Hierarchical Neural Network (HNN), that is able to deal with topological errors in the contour extraction. This strategy is extremely efficient for the majority of the classes: the recognition rate reaches about 99.5%. However, the performances sensitively decrease for 'similar characters', *i.e.* '8'/'B'. In that case, we adopt a strategy that revolves around decomposing the characters into structural elements. The Reeb graph generated from the binary images and a simple polygonal approximation permit to capture both topological and geometrical relevant features. The classification stage is carried out by a boosting algorithm.

Keywords: Fourier Descriptors, Hierarchical Neural Network, Reeb Graph, AdaBoost, Statistical-Structural Recognition

1 Introduction

We are here interested in the topic of character recognition from binary images of symbols, in the context of License Plate Recognition. The automatic identification of vehicles by the content of their license plate is an important research area that has been widely studied in the last two decades. A lot of industrial Car License Plate Recognition systems emerged since the eighties, and there today exist commercial OCRs at very affordable prices and impressive recognition rates (98-99%). However, these solutions implicitly make various assumptions that might be hard to fulfill in real environments or that make the system limited to some specific configurations. Actually, the development of reliable and robust systems for License Plate Recognition remains a very challenging task. The character recognition is the heart of the approach, and must be able to face various kinds of binary images corresponding to potentially degradated characters. For example, a limited resolution of the recorded characters in the image has to be managed, as well as dirty plates, screws and bolts, *etc.* The main challenge consists in finding a compromise between robustness and accuracy of the recognition. To fulfill this goal, we propose an hybrid strategy that combines advantages of statistical and structural approaches.

G. Bebis et al. (Eds.): ISVC 2007, Part II, LNCS 4842, pp. 529–538, 2007.

State of the Art. The character recognition part is the heart of the LPR system. It is composed of two main steps: feature extraction and feature classification. Concerning the methodology, we can distinguish two kinds of strategies: statistical and structural approaches. In statistical approaches, the input pattern is composed of a set of N features, and the classification stage can be seen as a statistical decision process. A general review of shape descriptors is beyond the scope of the paper and the reader can refer to [1]. For character recognition applications, the most basic feature that can be considered corresponds to the overall image extracted [2]. Semantically richer representations are commonly used: Fourier Descriptors [3], Curvature Scale Space [4], Zernike moments [5], Hu's moments [6], *etc.* Statistical methods strength is its robustness to noise. However, some small differences between similar patterns may be difficult to distinguish and therefore similar characters may be hard to recognize. On the other hand, structural approaches decompose the patterns into simpler primitives and obtain the properties of primitives and/or their relationships. Structural methods are closer to human recognition strategy [7], and are more powerful to recognize similar patterns. They are, however, more sensitive to noise. There are numerous approaches to decompose a given pattern into constitutive parts, but a general state of the art go beyond our goal. The methods of vectorization constitute an example that is closely related to our purpose. These techniques can be divided into several classes, varying with the final application of the method, and the background culture of the authors [8,9]. We will only focus on a few kinds of raster-to-vectors methodologies, largely developed for document analysis applications. In the run length based methods, the run length encoding (RLE) is a decomposition into elongated pixels along an axis of the image where we can build a line adjacency graph (LAG) [10]. The skeletonization and thinning methods are surely the most widely employed methods in vectorization. A survey of vectorization methods based on skeleton can be found in [11]. The aim is to compute a medial axis of the object that minimally represents its shape. The Hough transform is a line modelisation and is another common manner to detect lines and their intersections in an image [12]. Here, no relation is established between the detected lines, and thus the topology of the recognized object can not be deduced by this approach.

Contribution. We propose to use an hybrid approach for recognition, combining the advantages of statistical and structural methods. For the statistical recognition module (section 2.1), a hierarchical neural network analyzes the Fourier Descriptors extracted from the binary image. Similar to the system developed in [13], we notice that the recognition rate is very good for the majority of the classes, but largely decrease with 'similar' characters. In these cases, structural approaches are more adapted to discriminate the characters. Instead of detecting existing lines in the image, like in the LAG or with the Hough transform, we choose to first represent the topology of the represented objects, with the construction of the Reeb graph [14]. We also prefer not modifying the geometry of objects as the skeletonization process. Thus, we use an efficient vectorization algorithm, based on discrete geometry tools (section 2.2). Section 3 illustrates the approach efficiency by evaluating the performances of statistical and structural approaches. Finally section 4 concludes the paper and proposes directions for future works.

2 Proposed Approach

2.1 Statistical Character Recognition

2.1.1 The Fourier Descriptors

The Fourier Descriptors provide a discriminative signature of the contour of an object [15]. The N points of the contour $X(i) = \{x(i), y(i)\}, i \in \{1; N\}$ in the plane can be considered as complex number. The Fourier Descriptor $A(k)$ are determined by the computation of the Discrete Fourier Transform of the function X:

$$A(k) = \frac{1}{N} \sum_{i=0}^{N-1} X(i) e^{-j2\pi kn/N} \qquad , \; k \in \left\{ \frac{-N}{2}; \frac{N}{2} \right\} \tag{1}$$

The coefficients $A(k)$ represent the discrete contour of a shape in the frequency domain. The general shape of the object is represented by the lower frequency descriptors. Thus, it is possible to only keep a subset of the lowest frequency coefficients to extract a robust shape signature. Fourier Descriptors have intensively been used for shape matching applications and seem very adapted to our optical character recognition purpose. They prove to be superior to alternative representations, such as Autoregressive models [16], CSS or Zernike moments [17].

Normalization Process. Normalization is performed to provide a feature that only encodes shape. Invariance to translation and scale is performed by resizing the binary image. Moreover, we do not require rotation invariance in order to discriminate '6' from '9'. The set of Fourier coefficients is also dependent on the choice of the starting point. This is normalized in the frequency space. We use an approach similar to these proposed by Arbter et al [18]. After this normalization the starting point of the reconstructed contour is approximately at angle 0 (see [19]). Figure 1 illustrates the Fourier Descriptors computation and normalization.

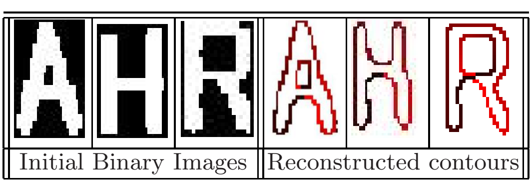

Fig. 1. Fourier Descriptors

2.1.2 The Classifier

The classification task is performed by means of a hierarchy of neural networks. The architecture is shown in Figure 2. At the first level, three different perceptrons, that we denote Cl_1, Cl_2 and Cl_3, are used. They are intended to recognize the external contour and the two possibly existing internal ones. The second level of the hierarchy takes as input the results provided by the 3 previous neural networks and carries out a final recognition, resulting to an output vector quantifying the probability of each possible character. We justify in 2.1.2.2 the chosen architecture. The classifier Cl_1 is dedicated to the outer contour analysis.

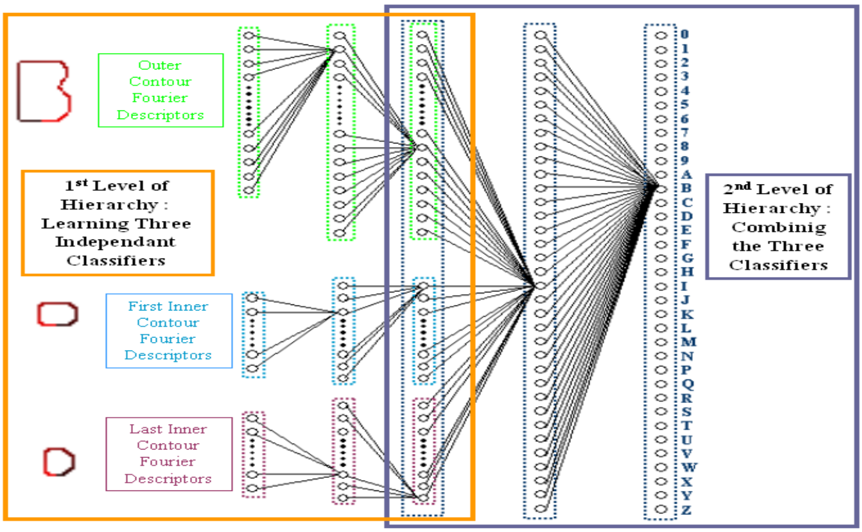

Fig. 2. Classifier for OCR Architecture

It includes the 35 different alphanumeric symbols with a single class for '0' and 'O', and is extended with topological error characters (see Figure 3). The Cl_2 classifier is devoted to identify the single inner contour of the '0', '4', 'A', 'D', 'Q', the upper contour of '8' and '9' (similar shape), and the upper contour of 'B', 'R' and 'P'. We introduce again classes with topological errors, corresponding to '8' and 'B' for which a single connected component has been extracted. Moreover a 'default' class encodes the absence of connected component. Finally, the Cl_3 classifier analyzes the shape of the lower inner contour of '8' and '6', and the lower contour of 'B'. The second level of the hierarchy precesses in a similar way as the first level. For a given input binary image, the classification output from Cl_1, Cl_2 and Cl_3 constitutes the input for the neural network.

2.1.2.1 Neural Network parameters. The internal architecture for each neural network is the following. The activation function for each neuron is a sigmoid one. The fundamental parameters of the neural network correspond to its weights, being iteratively updated during the training scheme. Similar to increasing number of recent works [20], we choose to develop an hybrid strategy, combining the accuracy of the gradient-based algorithms and the capacity of Genetic Algorithms to localize global minima. The evolutionary approach is used to initialize the weights for the backpropagation algorithm.

2.1.2.2 Discussion. We justify here the proposed architecture of the classifier and in particular its relevance for real applications. Decoupling the training between the 3 first level classifiers and the second level one brings us two major advantages. Firstly, the shape information inherent to each inner/outer contour is wholly capitalized on. For example, despite '6' and 'Z' differ each other from the presence of an inner contour, we want the recognition step to be able to discriminate them by means of the outer contour perceptron alone. Secondly, the hierarchy makes it possible to robustly learn characters with topological differences. For example, it is possible to properly identify alphanumeric with missing inner contours, as illustrated in Figure 3. These situations are actually common because the binary images come from an extraction step included in a global License Plate recognition framework, that is prone to failure.

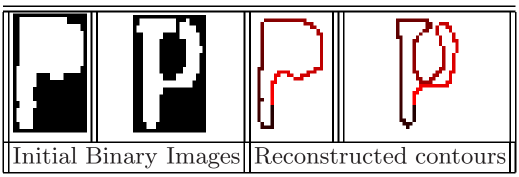

Fig. 3. Fourier Descriptors Sensitivity to topological differences

2.2 Structural Analysis

As we have presented in the previous section, we have chosen a structural approach based on discrete geometry tools to distinguish ambiguous characters in our system (*e.g.* '8'/'B') when the Neural Network classifier detects such a sample. We propose to describe the topology and the shape of such an example with an algorithm inspired from [21]. More exactely, this process performs the Reeb graph [14] of the binary object I and a simple polygonal structure inside I. For now, we deal with classifying '8'/'B' samples, but we could make a similar reasoning for others cases (*e.g.* '2'/'Z'). First, it can be noticed that those two characters really differ in the left part of the binary image I. Thus, we have decided to treat only this part of I, denoted I_L. As the letter is centered in the example, only the half of I has to be considered. We also choose an order to treat the pixels of I_L, *e.g.* from left to right and from top to bottom. Then, the algorithm `Struct_Analysis()` can be sketched as follows (see also Figure 4):

1. Group pixels of I_L to build as large as possible irregular pixels (or *cells*). We denote those k cells $\{C_i\}_{i=1,k}$. This part could be compared to a Run Length

Encoding scheme, where we have decided to merge together pixels with the same height. In an example 'B', the cells should be greater than in the '8' case. While this incremental process is performed, we build our topological and geometrical representation:

2. In the Reeb graph G_{I_L}, the nodes link connected components of I_L, which are -maybe thick- arcs of this object. Considering the order we have chosen before, a 'B' starts with one component, and finishes with three. So, in the general case, a 'B' example has one *begin* node in G_{I_L} (denoted by b in Figure 4), and three *end* nodes (e in this figure).
3. The polygonal structure P_{I_L} is computed by tracing the longest line segments through the arcs of I_L, according to the graph G_{I_L}. They respect the extended supercover model [22], which means that those line segments stand inside the recognized object. There may be more points in P_{I_L} if I_L represents an '8'.

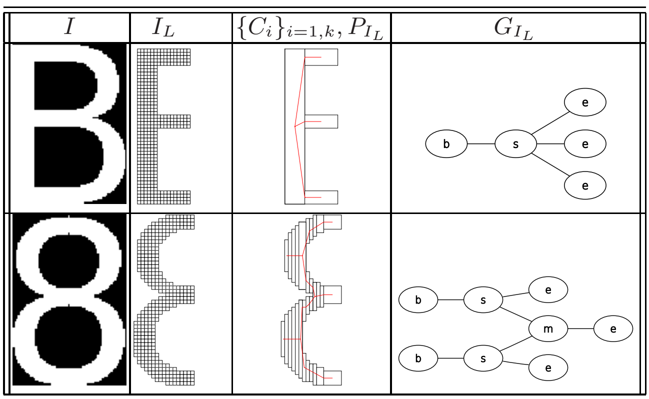

Fig. 4. The elements computed by the algorithm `Struct_Analysis(`I_L`)`. The Reeb graph G_{I_L} and the polygonal structure P_{I_L} are depicted for "perfect" letters '8' and 'B'. Nodes b and e are the beginning and the end of arcs in I_L, while s and m represent respectively splitting and merging operations on those arcs.

One could compare this process with classical Line Adjacency Graph (or LAG) based methods [10], but our final reconstruction with line segments contains naturally less points than in this kind of methodologies. Moreover, we can make a similar remark for the Reeb graph. Thus, we propose a fast algorithm, that do not need post-processing treatments and is convenient for real-time applications. For others letters, we would adapt this approach by treating another part of the binary image and thus by choosing an other orientation. For example, in the '2'/'Z' case, only the top part of the images mainly differs, and we choose a top-to-bottom orientation for `Struct_Analysis()`. In our system, we have implemented a structural classifier based on AdaBoost algorithm [23], since we have two classes C_1 and C_2 in our problem (*i.e.* $C_1 =' B'$, $C_2 =' 8'$).

Considering the remarks we have presented in the description of the algorithm `Struct_Analysis()`, let V be the feature vector defined as follows:

$$V^T = \begin{pmatrix} \#\{p = (x, y) \in P_{I_L}\} \\ \max_{i=1,k}\left(S(C_i)\right) \\ \#\{n = b, n \in G_{I_L}\} \end{pmatrix} \qquad (2)$$

Where $S(P_i)$ is the surface of the pixel P_i, and n is abusively denoted as a node *begin* in G_{I_L}. By giving a constant number N of examples per class, we first train the AdaBoost classifier with $2N$ feature vectors concatenated in a matrix $\mathbf{M}$. $\mathbf{M}$ is determined by a simple combinatorial algorithm based on the *Hill Climbing* [24] and permits to compute the best classification rate over the test sample. In the next section, we illustrate the performance of our structural approach, and the behavior of this optimization scheme.

3 Performances

Figure 5 presents the results of the statistical recognition. The performances are evaluated using a cross-validation procedure, each classifier being trained using 50 examples for each of its output class. In Figure 5-a, we can notice that the recognition rate is about 96%. More importantly, it must be pointed out that a large disparity exists among the different classes, the error rate for the similar characters ('8'/'B', '2'/'Z', '0'/'D'/'Q', and 'V'/'Y') being extremely higher than for the other classes. If we do not take into account these characters, the recognition rate reaches 99.5% (Figure 5-b). This justifies the validity of the hierarchical neural network to accurately discriminate the majority of the classes. We insist here on the fact that the binary images to classify come from a extraction and a binarization step. Thus, the topological differences after the contour extraction are common, and the hierarchical architecture makes it possible to robustly learn and recognize these variations. However, Figure 5-b illustrates as well the limitation of the Fourier Desciptors to discriminate similar pattern. Indeed, these classes only differ locally and are distinguishable by sailent features such as right angles that are lost after the lowest frequency terms extraction.

In Figure 6, we show an example of the behaviour of our combinatorial optimization scheme during 10 000 iterations (about 2 hours). This experiment is performed over a base of about 446 '8' and 'B' images, and we have fixed $N = 50$. Then, the AdaBoost classifier learns the best-rate base ($2N = 100$ images) and runs the classification over the 336 other images. After a few iterations, the global rate is more than the one obtained by the neural network approach (about 85%) and increases until more than 90% after a hundred iterations (1.5 minutes). The best score (more than 93%, see Table 1) is obtained in 25 minutes in this example. Any user may choose the quality of the structural classification and have a coarse result (but better than the neural network) in a few seconds. It could also let the algorithm run until he gets an accurate enough classification.

In Table 1, we depict the best rate of structural classification for the same base of 'B' and '8' images. We can notice that the recognition rate is better by

Classes	Number of Errors	Number of Examples	Error Rate
0	25	248	10,08%
1	0	241	0,00%
2	4	253	1,58%
3	3	290	1,03%
4	3	175	1,71%
5	0	269	0,00%
6	6	384	1,56%
7	1	261	0,38%
8	98	410	23,90%
9	1	335	0,30%
A	1	191	0,52%
B	35	276	12,68%
C	0	95	0,00%
D	15	156	9,62%
E	0	63	0,00%
F	0	55	0,00%
G	4	140	2,86%
H	0	76	0,00%
I	1	56	1,79%
J	0	78	0,00%
K	0	134	0,00%
L	0	85	0,00%
M	0	103	0,00%
N	0	88	0,00%
P	0	183	0,00%
Q	5	59	8,47%
R	0	158	0,00%
S	0	131	0,00%
T	0	136	0,00%
U	0	53	0,00%
V	6	76	7,89%
W	0	72	0,00%
X	0	78	0,00%
Y	0	61	0,00%
Z	0	123	0,00%
Total	208	5592	3,72%

a) Performances for all classes

Classes	Number of Errors	Number of Examples	Error Rate
1	0	241	0,00%
3	3	290	1,03%
4	3	175	1,71%
5	0	269	0,00%
6	6	384	1,56%
7	1	261	0,38%
9	1	335	0,30%
A	1	191	0,52%
C	0	95	0,00%
E	0	63	0,00%
F	0	55	0,00%
G	4	140	2,86%
H	0	76	0,00%
I	1	56	1,79%
J	0	78	0,00%
K	0	134	0,00%
L	0	85	0,00%
M	0	103	0,00%
N	0	88	0,00%
P	0	183	0,00%
R	0	158	0,00%
S	0	131	0,00%
T	0	136	0,00%
U	0	53	0,00%
W	0	72	0,00%
X	0	78	0,00%
Total	20	3930	0,51%

b) Performances without similar pattern

Fig. 5. Global Performances of the Statistical Classifier

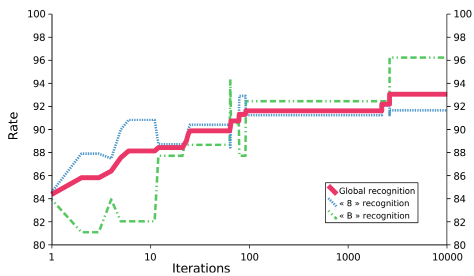

Fig. 6. An example of the run of our classifier optimization algorithm during 10 000 iterations (about 2 hours). Individual '8' and 'B' recognition rates are depicted in dashed lines, and the global rate in solid lines

Table 1. Recognition rates, depending on the method used and the set of images treated

	Neural Network	Structural classification
'8' images	76.1 %	92.9 %
'B' images	86.96 %	95.28 %
Global	80.61 %	**93.64 %**

considering the structural classifier, for the sets of 8 and B images. This rate increases by 13% for all the image base, and reaches more than 93%.

4 Conclusion and Future Works

We propose an original hybrid approach for Optical Character Recognition in the context of a License Plate Recognition system. A statistical method analyzes the extracted Fourier Descriptors with a Hierarchical Neural Network. Decoupling the two levels of the Hierarchy for the training makes the approach robust to noise, over-fitting or offers the capacity to properly identify alphanumeric symbols with topological errors. The statistical recognition performances are impressively good for the majority of the classes, but sensitively decrease for 'similar' characters. The lack of locality of the Fourier Descriptors and the high frequency terms that are ignored during the feature extraction step constitute the origin of the problem. Thus, we use for these characters a structural approach that extracts discriminative local features to perform the recognition. Thanks to discrete geometrical tools, we represent topological and geometrical important features of the treated binary images. In this article, the error rate for '8'/'B' is shown to be reduced of more than 13%. The perspective for future works would be to validate the structural approach for the other classes of similar pattern.

References

1. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1) (2004)
2. Bremananth, R., Chitra, A.: A robust video based license plate recognition system. ICISIP (2005)
3. Shridhar, M., Badreldi, A.: High accuracy character recognition algorithm using fourier and topological descriptors. Patt. Recogn. 17(5), 515–524 (1984)
4. Kopf, S., Haenselmann, T., Effelsberg, W.: Enhancing Curvature Scale Space Features for Robust Shape Classification. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2005), Amsterdam, The Netherlands (2005)
5. Broumandnia, A., Shanbehzadeh, J.: Fast zernike wavelet moments for farsi character recognition. Image Vision Comput. 25(5), 717–726 (2007)
6. Cash, G., Hatamian, A.: Optical character recognition by the method of moments. 39(3), 291–310 (1987)
7. Pavlidis, T.: 36 years on the pattern recognition front. Pattern Recognition Letters 24, 1–7 (2003)
8. Cordella, L.P., Vento, M.: Symbol recognition in documents: a collection of techniques? IJDAR 3(2), 73–88 (2000)
9. Hilaire, X., Tombre, K.: Robust and accurate vectorization of line drawings. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(6), 890–904 (2006)
10. Burge, M., Kropatsch, W.: A minimal line property preserving representation of line images. Computing 62(4), 355–368 (1999)
11. Lam, L., Lee, S.W., Suen, C.Y.: Thinning methodologies - a comprehensive survey, 61–77 (1995)

12. Kälviäinen, H., Hirvonen, P., Xu, L., Oja, E.: Probabilistic and non-probabilistic hough transforms: overview and comparisons. Image Vision Comput. 13(4), 239–252 (1995)
13. Pan, X., Ye, X., Zhang, S.: A hybrid method for robust car plate character recognition. Engineering Applications of Artificial Intelligence, 963–972 (2005)
14. Reeb, G.: Sur les points singuliers d'une forme de pfaff complément intégrable ou d'une fonction numérique. Comptes Rendus de L'Académie ses Séances, 847–849 (1946)
15. Peerson, E., Fu, K.: Shape discrimination using fourier descriptor. Trans. Syst. Man, Cybern. SMC-7(3), 170–179 (1977)
16. Kauppinen, H., Seppanen, T., Pietikainen, M.: An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2), 201–207 (1995)
17. Zhang, D., Lu, G.: Content-based shape retrieval using different shape descriptors: A comparative study. icme 00, 289 (2001)
18. Arbter, K., Snyder, W.E., Burhardt, H., Hirzinger, G.: Application of affine-invariant fourier descriptors to recognition of 3-d objects. IEEE Trans. Pattern Anal. Mach. Intell. 12(7), 640–647 (1990)
19. Folkers, A., Samet, H.: Content-based image retrieval using fourier descriptors on a logo database. In: ICPR 2002: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR 2002), vol. 3, IEEE Computer Society, Washington, DC, USA (2002)
20. Devi, S., Panda, S., Shyam, S., Pattnik, D., Khuntia, B.: Initializing artificial neural networks by genetic algorithm to calculate the resonant frequency of single shorting post rectangular patch antenna. In: IEEE Antennas and Propagation Society International Symposium, pp. 144–147 (2003)
21. Vacavant, A., Coeurjolly, D., Tougne, L.: Topological and geometrical reconstruction of complex objects on irregular isothetic grids. In: Kuba, A., Nyúl, L.G., Palágyi, K. (eds.) DGCI 2006. LNCS, vol. 4245, pp. 470–481. Springer, Heidelberg (2006)
22. Coeurjolly, D.: Supercover model and digital straight line recognition on irregular isothetic grids. In: Andrès, É., Damiand, G., Lienhardt, P. (eds.) DGCI 2005. LNCS, vol. 3429, pp. 311–322. Springer, Heidelberg (2005)
23. Schapire, R.: The boosting approach to machine learning: An overview (2001)
24. Selman, B., Levesque, H.J., Mitchell, D.G.: A new method for solving hard satisfiability problems. In: AAAI, pp. 440–446 (1992)