

Report on the Third Contest on Symbol Recognition

Ernest Valveny¹, Philippe Dosch², Alicia Fornes¹ and Sergio Escalera¹

¹ Computer Vision Center, Dep. Ciències de la Computació
Universitat Autònoma de Barcelona, Bellaterra (Spain)

{`ernest,afornes,sergio.escalera@cvc.uab.es`}

² LORIA, Université Nancy 2, Nancy, France

{`Philippe.Dosch@loria.fr`}

Abstract. In this paper we make a brief report of the third edition of the International Symbol Recognition Contest, organized in the context of GREC'07. This contest follows the series started at the GREC'03 workshop. In this report we describe the main changes introduced in the test data according to the conclusions of the past edition of the contest. We also summarize the results obtained by the only participant method. Finally, we point out some conclusions and open issues to be addressed in the next editions of the contest.

1 Introduction

The performance evaluation of symbol recognition has been a focus of research interest in the last years. Several surveys on symbol recognition [1–4] pointed out the need of standard evaluation tools in order to compare the large number of symbol recognition methods. As a result, a generic framework for the evaluation of symbol recognition has been proposed [5]. In this framework, the main issues to be addressed by any performance evaluation system are identified (mainly, the generation of datasets and groundtruth, the definition of metrics, and the protocol of evaluation) and several alternatives are proposed and discussed in the special case of symbol recognition.

Following this generic framework, and from a practical point of view, several contests have been organized. Actually, the first effort on the evaluation of symbol recognition was undertaken at ICPR'00 [6] where a contest was proposed using a dataset consisting of 25 electrical symbols, which were scaled and degraded with a small amount of binary noise. Afterwards, the series of contests on symbol recognition in the context of the GREC workshop started in 2003. In the first edition [7], the dataset was composed of 50 architectural and electrical symbols. These symbols were rotated, scaled, degraded with binary noise and deformed through vectorial distortion in order to generate up to 72 different tests with increasing levels of difficulty and number of symbols. There were five methods participating in the contest. Then, in the second edition [8] some modifications were introduced according to the conclusions of the first contest.

The set of symbols was increased up to 150 different symbols, allowing the definition of more pertinent tests for the evaluation of the scalability. In addition, four new degradation models were added to the framework for the generation of more noisy data. These new degradation models constituted a kind of "torture models". In this way, the robustness of the methods could be tested under very extreme conditions. Four methods participated in the contest.

Among the main conclusions stated in the report of the last contest [8] we can remark some issues that have been taken into account, not only in the design of the third edition of the contest, but also in the work undertaken in the last two years. Firstly, it was stated that evaluation should be a continuous task, not concentrated every two years at specific contests. Therefore, tools for the analysis of the results of recognition methods should be provided. In this sense, the work on the French project ÉPEIRES³ has set up a web-based framework for the evaluation of symbol recognition where new tests can be easily created and the results obtained by a given method can be uploaded and automatically analyzed. Secondly, it was stressed the need of extending the evaluation to symbol localization and segmentation. Some work on this topic has been undertaken under the framework of the ÉPEIRES project too. As a result, a first approach to the generation of synthetic complete architectural drawings has been developed[9]. This is the first step in order to be able to generate large amounts of data for the evaluation of segmentation. Work has still to be done concerning the metrics to compare the results with the ground-truth. So, in the third edition of the contest we have not considered localization and segmentation and we have constrained the contest only to pre-segmented symbols as in past editions. Thirdly, it was claimed that more heterogeneous data should be included in the framework. In order to give an answer to this demand, we have included in this edition of the contest a dataset composed of logos. Logos are also graphic symbols, but with very different properties (regarding shape, primitives, appearance, etc) with respect to the technical symbols used in the previous contests. In this way, the range and variability of symbols is extended. Finally, it was remarked the need of defining blind tests in order to ensure that participant methods are not adapted to the particular data of the contest. In this edition this remark has been taken into account by including different types of randomly selected degradations in the same test. The goal is to be sure that participants design generic symbol recognition methods, able to work with all kind of (noisy) symbols.

In the next sections, we describe more in details the data provided in this edition of the contest as well as the results obtained by the only participant method. But before, we would like to recall the original purpose of this series of contests as stated in the call for participation: the main goal is not to give a single performance measure for each method, but to provide a tool to compare various symbol recognition methods under several different criteria. The question consists of determining the performance of symbol recognition methods when working on various kinds of symbols, extracted from diverse application domains, under several constraints, with different levels of noise and degradation.

³ <http://www.epeires.org/>

Whatever the performance measures are, we strongly believe that the main objective of this evaluation framework must be the scientific analysis of the results. This analysis must be intended to determine the different qualities expected for recognition methods: robustness, genericity, precision, computational efficiency.

The paper is organized as follows. In section 2 we describe the datasets that were generated for this edition of the contest. Then, in section 3 we briefly describe the main features of the only participant method and analyze the results of its application to the dataset. Finally, in section 4 we state the main conclusions of the evaluation and some actions to be undertaken in the future.

2 Dataset

As explained in the previous section, we have considered two different kinds of symbols in this edition of the contest: technical symbols and logos. For technical symbols, we used the same dataset as in the last edition, that is, a set of 150 symbols, mainly originally from the domains of architecture and electronics. We can see in figure 1 some examples of this dataset where symbols are composed of linear primitives (straight lines and arcs). Logos are the main novelty in the dataset. We have included them in order to extend the spectrum of symbols. Logos are different of technical symbols in the sense that they are not composed only of linear primitives. They can include solid regions, texture, characters, more than one graphic component, etc. Thus, it is a completely different kind of symbol representation and can be useful to test whether recognition methods are generic enough. This dataset is composed of 105 different logos and some examples can be seen in figure 2.

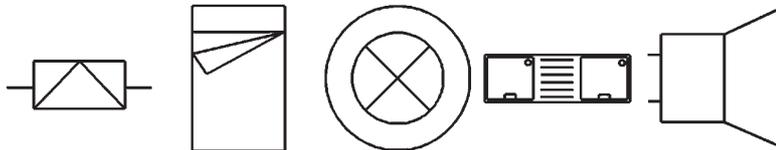


Fig. 1. Some examples of technical symbols.

We have used the same kind of transformations and degradations as in the last contest to generate the final tests for evaluation. Thus, rotation, scaling and binary degradation using the Kanungo's method [10] have been applied to the ideal models of the symbols. In figure 3, we can see some examples of the degraded images. We have considered the same six models of degradation defined in the last contest as it was concluded that no new models were needed. As explained in the previous section, some of these models introduce heavy distortions in the images and thus, the level of difficulty is high.



Fig. 2. Some examples of logos.



Fig. 3. Some examples of degraded images.

The final tests for the evaluation have been generated combining all these elements. In table 1, there is a summary of all the tests with their main features. We can see that we have designed tests for two different sizes of the database for technical symbols. A first set of tests with 50 symbols and a second set with 150 symbols. In this way, we can evaluate the robustness to the scalability in the number of symbols. For both sets, all the possible combinations of rotation and scaling have been considered. Moreover, all the tests include binary degradation. Degradation is always randomly selected among the six possible models. Thus, we achieve the goal of generating blind tests, as explained in the introduction.

For logos, all the tests include the whole database of 105 symbols. In this case, several combinations of rotation, scaling and degradation have been considered. Two tests including specific models of degradation have been defined but, for the rest of the tests, degradation is randomly selected in order to generate blind tests.

All the information and data related to the tests can be found on the webpage of the ÉPEIRES project at <http://www.epeires.org/>.

3 Results

In this edition, only one method participated in the evaluation of the proposed tests. The method has been developed by Alicia Fornes and Sergio Escalera, from the Computer Vision Center, in Spain. A paper describing this method appears in the current LNCS volume. Nevertheless, we give an overview of the method in the next section in order to facilitate the understanding of the results.

| Test | Dataset | No. of Models | No. of Images | Rotation | Scaling | Degradation |
|------|-----------|---------------|---------------|----------|---------|-------------------------------|
| 1 | Technical | 150 | 500 | Random | None | Random among 6 GREC'05 models |
| 2 | Technical | 150 | 500 | None | Random | Random among 6 GREC'05 models |
| 3 | Technical | 150 | 500 | Random | Random | Random among 6 GREC'05 models |
| 4 | Technical | 50 | 200 | Random | None | Random among 6 GREC'05 models |
| 5 | Technical | 50 | 200 | None | Random | Random among 6 GREC'05 models |
| 6 | Technical | 50 | 200 | Random | Random | Random among 6 GREC'05 models |
| 7 | Logos | 105 | 300 | Random | None | None |
| 8 | Logos | 105 | 300 | None | Random | None |
| 9 | Logos | 105 | 300 | Random | Random | None |
| 10 | Logos | 105 | 300 | None | None | Second GREC'05 model |
| 11 | Logos | 105 | 200 | None | None | Fourth GREC'05 model |
| 12 | Logos | 105 | 300 | None | None | Random among 6 GREC'05 models |
| 13 | Logos | 105 | 300 | Random | None | Random among 6 GREC'05 models |
| 15 | Logos | 105 | 200 | Random | Random | Random among 6 GREC'05 models |

Table 1. Description of all the tests.

3.1 Description of the method

The method works on the skeleton or the contour of the original image. The choice use of skeletons or contours is decided depending on the shape database. Skeletons are preferred for line-based symbols while contours are dedicated for silhouette-based shapes. Images are aligned using the Hotelling transform that is based on principal components to find the main axis of the object. Then, the shape is represented using the Blurred shape model descriptor (BSM) that makes the technique robust against elastic deformations. Afterwards, Adaboost is applied to each pair of classes to train a set of binary classifiers. Finally, the set of binary classifiers is embedded in the framework of Error Correcting Output Codes (ECOC) to improve the final classification.

The main core of this method is the BSM descriptor. With this descriptor, the symbol is described by a probability density function that encodes the probability of pixel densities of image regions: The image is divided in a grid of $n \times n$ equal-sized subregions. Every bin receives votes from the pixels in its region but also from the pixels in the neighboring bins. The weight of the vote is set according to the distance to the center of the bin. The output descriptor is a vector histogram where every position corresponds to the weight of the pixels in the context of every sub-region. This vector is normalized in the range $[0..1]$ to obtain the probability density function (pdf) of the $n \times n$ bins. In this way, the output descriptor represents a distribution of probabilities of the object shape considering spatial distortions. For further details, see [11].

3.2 Analysis of results

Unfortunately we cannot present results for all the tests. The participant method was only evaluated using 5 of the proposed tests. In table 2, we show the recognition rates of the method for these 5 tests.

| Test | Dataset | Rotation | Scaling | Degradation | Recognition rate |
|------|-----------|----------|---------|--------------|------------------|
| 5 | Technical | None | Random | Random | 91% |
| 8 | Logos | None | Random | None | 95% |
| 10 | Logos | None | None | Second model | 82% |
| 11 | Logos | None | None | Fourth model | 46% |
| 12 | Logos | None | None | Random | 55% |

Table 2. Results of the method.

If we try to analyze these results we can draw several conclusions. Only one test with technical symbols was evaluated. This test contains images of 50 symbols with scaling and binary degradation. The recognition rate, 91%, can be considered as a good result if we compare it with the recognition rates obtained for similar tests in the past contest. In it, the average of the recognition rates for all the methods, all degradation models and scaling was only 74.25%.

Concerning logos, the recognition rate for images without degradations remains at a high level, 95%. However, it decreases rapidly when degradations are applied. Although we have no other methods to compare these results, we can try to establish some relations with the results obtained in the most similar kind of tests in the last contest. In that case, for tests with 100 symbols (approximately the same number of logos), no scaling and binary degradation, the average of all the methods over all models of degradation was 90%, clearly greater than the recognition rate obtained in this case for the test 12 with logos. It is difficult to draw exact conclusions from these results as we have no other results with the logo database. We cannot state whether the low results for the logos are due to the fact that logos are intrinsically more difficult to recognize than technical symbol or whether they are a consequence that this method is better adapted to linear shapes than to solid shapes.

4 Conclusions and future work

In this edition, we have extended the contest with two of the considerations arising from the conclusions of the last contest: we have included a new kind of symbols, logos, and we have generated blind tests combining all the models of degradations. However, no relevant conclusions can be drawn from the experimentation with the logo dataset as we only have results from one method, and not for all the tests.

Nevertheless, after three editions of the contest, the framework for the evaluation of the recognition of pre-segmented symbols recognition seems mature enough. In this sense, this framework can be converted in a tool for continuous evaluation through the web platform of the ÉPEIRES project. This way, any researcher can contribute with new results to the database of the platform and we can have a good overview of the performance of a large number of methods. In this context, many tests have been generated along the three editions of the contest. Maybe it would be interesting to define a set of standard validation tests taking into account all the kinds of transformations and degradations. This set would constitute a kind of standard evaluation that every method should pass. Thus, we would have a generic global evaluation of all the methods. In addition, it would be also interesting to add new symbols to the framework in order to create a really large database of symbols, representative enough of all kinds of graphic symbols.

The big challenge that is still to be addressed is the evaluation of localization/segmentation in complete drawings with non-segmented symbols. In this sense, some advances have been described in the field of ground-truthing with the generation of synthetic documents. The next step should be the definition of metrics to compare the results with the ground-truth, and the definition of the evaluation protocol. We plan to advance in this direction and we hope to be able to propose early a contest on symbol localization.

Finally, we want to make a note on the low participation in this edition of the contest. For next editions, we should increase the efforts in order to promote the participation in the contest. However, this could be another point for providing a continuous framework for the evaluation of the recognition of pre-segmented symbols. We hope that new researchers will be interested by the contest when it will include symbol localization.

Acknowledgment

The authors would like to acknowledge the French Ministry of Research for the funding of the ÉPEIRES project as a part of the Techno-Vision campaign.

This work has also been partially supported by the Spanish project TIN2006-15694-C02-02, and by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018)

References

1. Chhabra, A.K.: Graphic Symbol Recognition: An Overview. In Tombre, K., Chhabra, A.K., eds.: Graphics Recognition—Algorithms and Systems. Volume 1389 of Lecture Notes in Computer Science. Springer-Verlag (1998) 68–79
2. Cordella, L., Vento, M.: Symbol recognition in documents : a collection of techniques. International Journal on Document Analysis and Recognition (IJ DAR) **3** (2000) 73–88 DIA.

3. Lladós, J., Valveny, E., Sánchez, G., Martí, E.: Symbol recognition: Current advances and perspectives. In Blostein, D., Kwon, Y., eds.: *Graphics Recognition: Algorithms and Applications, Selected Papers from Fourth International Workshop on Graphics Recognition, GREC'01*. Springer, Berlin (2002) 104–127 Volume 2390 of *Lecture Notes in Computer Science*.
4. Tombre, K., Tabbone, S., Dosch, P.: Musings on symbol recognition. In: *Workshop on Graphics Recognition (GREC)*. Volume 3926 of *Lecture Notes in Computer Science (LNCS)*. (2005) 23–34 DIA.
5. Valveny, E., al.: A general framework for the evaluation of symbol recognition methods. *International Journal on Document Analysis and Recognition (IJ DAR)* **1** (2007) 59–74 Performance Evaluation.
6. Aksoy, S., Ye, M., Schauf, M., Song, M., Wang, Y., Haralick, R., Parker, J., Pivovarov, J., Royko, D., Sun, C., Farneboock, G.: Algorithm performance contest. In: *Proceedings of 15th. International Conference on Pattern Recognition*. Volume 4. (2000) 870–876 Barcelona, Spain.
7. Valveny, E., Dosch, P.: Symbol recognition contest: a synthesis. In Lladós, J., Kwon, Y.B., eds.: *Graphics Recognition: Recent Advances and Perspectives – Selected papers from GREC'03*. Volume 3088 of *Lecture Notes in Computer Science*. Springer-Verlag (2004) 368–385
8. Dosch, P., Valveny, E.: Report on the second symbol recognition contest. In: *Workshop on Graphics Recognition (GREC)*. Volume 3926 of *Lecture Notes in Computer Science (LNCS)*. (2006) 381–397 Performance Evaluation.
9. Delalandre, M., Pridmore, T., Valveny, E., Trupin, E., Locteau, H.: Building synthetic graphical documents for performance evaluation. In: *Workshop on Graphics Recognition (GREC)*. (2007) 84–87
10. Kanungo, T., Haralick, R.M., Baird, H.S., Stuetzle, W., Madigan, D.: Document Degradation Models: Parameter Estimation and Model Validation. In: *Proceedings of IAPR Workshop on Machine Vision Applications, Kawasaki (Japan)*. (1994) 552–557
11. Fornés, A., Escalera, S., LLadós, J., Sánchez, G., Radeva, P., Pujol, O.: Handwritten symbol recognition by a boosted blurred shape model with error correction. In: *3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2007)*. Volume 4477 of *Lecture Notes in Computer Science*. Springer-Verlag (2007) 13–21