Report on the Symbol Recognition and Spotting Contest

Ernest Valveny[1], Mathieu Delalandre[2], Romain Raveaux[3], and Bart Lamiroy[4]

[1] Computer Vision Center, Univ. Autònoma de Barcelona
Edifici O, Campus UAB, 08193 Bellaterra, Spain
ernest@cvc.uab.es

[2] Laboratoire d'Informatique
Université de Tours, France
mathieu.delalandre@univ-tours.fr

[3] L3i laboratory
Université de La Rochelle, France
romain.raveaux01@univ-lr.fr

[4] LORIA / INPL - École des Mines de Nancy
Bart.Lamiroy@loria.fr

Abstract. In this paper we summarize the framework and the results of the fourth edition of the International Symbol Recognition Contest, organized in the context of GREC'11. The contest follows the series started at the GREC'03 workshop and it is the first time that, in addition to recognition of isolated symbols, the contest includes the evaluation of symbol spotting. In this report we describe the evaluation framework – including datasets and evaluation measures – and we summarize the results obtained by the only participant method.

Keywords: Performance evaluation, symbol recognition, symbol spotting.

1 Introduction

Symbol recognition has been a topic of active research within the graphics recognition community with many different approaches described in the literature [2,7,11]. Thus, there is a real need for a generic and standard framework that permits a fair comparison of all existing methods. Such a framework was discussed in [12] in terms of datasets, ground-truth, evaluation metrics and protocol of evaluation. Following these ideas, several competitions have been organized. The first work on evaluation of symbol recognition was undertaken at ICPR'00 [1]. The dataset consisted of 25 electrical symbols that were scaled and degraded with a small amount of binary noise to generate images of non-connected symbols. The series of contests on symbol recognition linked to the GREC workshop started in 2003. In the first edition [13], the dataset was composed of 50 architectural and electrical symbols that were rotated, scaled, degraded with binary noise and deformed through vectorial distortion in order to

Y.-B. Kwon and J.-M. Ogier (Eds.): GREC, LNCS 7423, pp. 198–207, 2013.

generate up to 72 different tests with increasing levels of difficulty and number of symbols. In the second edition of the contest [4] the set of symbols was increased up to 150 different symbols, allowing the definition of more pertinent tests for the evaluation of the scalability. Degradation models included some very extremely hard models in order to test the robustness of the methods under very extreme conditions. In the third edition [5] a dataset of logos was included in the framework in order to test the genericity of the participant methods. With the same goal different types of randomly selected degradations were included in the same test in order to generate blind tests.

In this paper we summarize the framework and the results of the new edition of the contest following the series of previous GREC contests. Three are the main novelties of this edition of the contest. Firstly, a new set of images for isolated symbol recognition is generated. This new set is composed of a set of blind tests – mixing different kinds of degradations in the same test – and intends to be representative enough of the kind of degradations encountered in graphics recognition applications. It has been carefully designed to permit the evaluation of the scalability of the methods. Secondly, a new type of test has been created including images of symbols that have been directly cropped from real drawings. The goal is to evaluate the performance of isolated symbol recognition when it is not possible to achieve a perfect segmentation of the symbol. Thirdly, a set of complete architectural and electrical drawings has been defined allowing to include, for the first time, the evaluation of symbol spotting. This was one of the missing issues in the past editions of the contest. Recently, there have been interesting contributions regarding both the creation of datasets [3] and the definition of metrics [10] for performance evaluation of spotting systems in graphics recognition. We have taken advantage of these works to include symbol spotting in this edition of the contest.

In the rest of the paper, in section 2 we describe the datasets generated for the contest. Then, in section 3 we explain the evaluation metrics used both for recognition and spotting. In section 4 we analyze the results obtained by the only participant method. Finally, in section 5 we state the main conclusions and discuss open issues for next editions of the contest.

2 Dataset

For the generation of the dataset we have used the same set of 150 symbols of the previous GREC contests. We have created different datasets for symbol recognition and symbol spotting that are described in the next sections. Tables 1 and 2 summarize the contents of these datasets for training and test respectively.

2.1 Symbol Recognition

Datasets for isolated symbol recognition have already been generated for the past editions of the contest. However, we decided to create new datasets in order to provide a set of tests that could complement some of the drawbacks of previous ones and could become a kind of generic datasets to be used from now on as a reference for any evaluation of symbol recognition methods. Thus, we designed the new datasets with the following goals: first, to provide a set of tests that could evaluate scalability of methods; second, to be able to test the performance of methods under some realistic increasing degradations; third, to be able to test the genericity of the methods. It is worth to mention that this option prevents from comparing the results with those of previous contests.

Table 1. Training datasets. (S/M is the instance of (S)ymbols per (M)odel)

id	**Type**	**Domain**	**Models**	**S/M**	**Symbols**	**Noise**
#1	recognition	Technical	150	10	1500	Rotation
#2	recognition	Technical	150	10	1500	Scaling
#3	recognition	Technical	150	10	1500	Rotation/Scaling
#4	recognition	Technical	150	25	3750	Kanungo-Level α
#5	recognition	Technical	150	25	3750	Kanungo-Level β
#6	recognition	Technical	150	25	3750	Kanungo-Level η
#7	recognition	Technical	36	25	900	Context
					16650	
id	**Type**	**Domain**	**Models**	**Images**	**Symbols**	**Noise**
#8	localization	Architectural	16	5	142	Ideal
#9	localization	Architectural	16	5	133	Kanungo-Level 1
#10	localization	Architectural	16	5	144	Kanungo-Level 2
#11	localization	Architectural	16	5	128	Kanungo-Level 3
#12	localization	Electrical	21	5	54	Ideal
#13	localization	Electrical	21	5	81	Kanungo-Level 1
#14	localization	Electrical	21	5	91	Kanungo-Level 2
#15	localization	Electrical	21	5	62	Kanungo-Level 3
				40	**835**	

As a result we generated three different sets of images each with an increasing number of symbols (50, 100 and 150). For each of these tests, we synthetically generated 50 images of every symbol with different degradations. To generate degradations, as in previous contests, we used the method of binary degradation proposed by Kanungo et al.[6]. This is a well founded and established method for generating document distortions. We started by generating basic images of each symbol by applying a very slight binary degradations to the ideal image of the symbol – figure 1(a)–. Using these basic images we generated a set of images with rotation, scaling and combined rotation and scaling (training sets #1 to #3 in table 1). Then, we generated more degraded images according to different settings of Kanungo's method parameters. We just modified each of the parameters independently in order to get a set of increasing different types of distortions. Changing parameter α we were able to generate images where lines are thinned with respect to the original ones – figure 1(b)-(c) and set #4 in table 1–. Parameter β allows to simulate thicker lines – figure 1(d)-(e) and set #5 in table 1–. Finally, parameter η influences the level of global noise – figure 1(f)-(g) and set #5 in table 1–. In order to test the genericity of methods we mixed randomly all degradations in the final tests so that participants couldn't have any a priori information about the kind of noise of images – see table 2–.

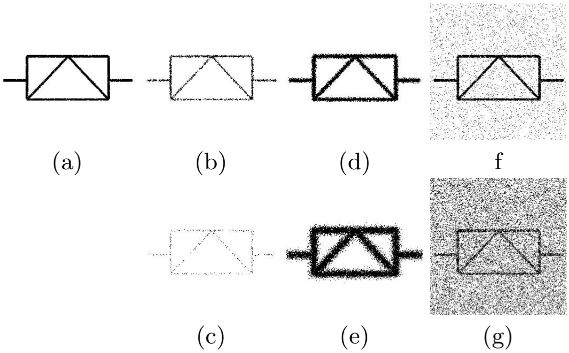

Fig. 1. Examples of images generated for the symbol recognition tests. (a) Basic image. (b)-(c) Degradation according to parameter α. (d)-(e) Degradation according to parameter β. (f)-(g) Degradation according to parameter η.

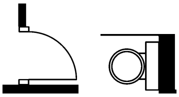

Fig. 2. Examples of images cropped from complete drawings

In addition to these three sets of images of isolated symbols, we generated an additional fourth set consisting of images directly cropped from complete drawings. Thus, images were instances of symbols not perfectly segmented. The goal of this fourth test set was to evaluate recognition performance under more realistic conditions where a perfect segmentation can not be usually achieved. It can only be seen as a way of involving user interaction in the tests. These tests propose query symbols (i.e. cropped images of symbols) that can be affected by the way the user makes the selection. They try to imitate this effect by randomly growing the bounding box of the symbol. In that sense, they constitute a tradeoff between the recognition and localization problems. This work has been motivated by the interest of the community on such a problem, as highlighted in some recent contributions [8]. Only 36 different symbols were used to generate this set. Some examples of these images can be seen in figure 2 – tests #7 and #4 in tables 1 and 2 respectively –.

2.2 Symbol Spotting

This is the first time that images of complete drawings are provided for evaluation of symbol spotting in the series of GREC contests. The main difficulty up to now was the unavailability of public datasests for symbol spotting. In this edition we have taken advantage of a recent work describing the synthetic generation of complete architectural floorplans and electronic diagrams [3]. The approach is based on the definition of a set of constraints that directs the placement of a

Table 2. Final datasets. (S/M is the instance of (S)ymbols per (M)odel)

id	Type	Domain	Models	S/M	Symbols	Noise
#1	recognition	Technical	50	50	2500	Kanungo-Mixed
#2	recognition	Technical	100	50	5000	Kanungo-Mixed
#3	recognition	Technical	150	50	7500	Kanungo-Mixed
#4	recognition	Technical	36	50	1800	Context
					16800	
id	**Type**	**Domain**	**Models**	**Images**	**Symbols**	**Noise**
#5	localization	Architectural	16	20	633	Ideal
#6	localization	Architectural	16	20	597	Kanungo-Level 1
#7	localization	Architectural	16	20	561	Kanungo-Level 2
#8	localization	Architectural	16	20	593	Kanungo-Level 3
#9	localization	Electrical	21	20	246	Ideal
#10	localization	Electrical	21	20	274	Kanungo-Level 1
#11	localization	Electrical	21	20	237	Kanungo-Level 2
#12	localization	Electrical	21	20	322	Kanungo-Level 3
				160	**3463**	

given set of symbols on a pre-defined background according to the properties of a particular domain (architecture, electronics, engineering, etc.). In this way, we can obtain a large amount of images resembling real documents by simply defining the set of constraints and providing a few pre-defined backgrounds. As documents are synthetically generated, the groundtruth (the location and the label of every symbol) becomes automatically available. All the documents generated in the context of this work have been published in a dataset called SESYD[1] made publicly available[2] for performance evaluation purpose.

To generate the localization tests for this GREC contest, we have used samples of the SESYD dataset. The whole SESYD dataset is composed of 20 collections, 10 collections from the architectural domain plus 10 from the electrical one. The architectural floorplans are composed of 16 symbol models whereas the electrical diagrams are composed of 21. We have selected 14 collections from the initial dataset, those that permit to guarantee the homogeneity of line thickness across different images. Images have been randomly selected in order to get a mix of different backgrounds in tests . Tables 1, 2 give the details about these tests. We have generated 8 different tests both for training (#8 to #15) and test (#5 to #12) datasets, four corresponding to architectural floorplans and four corresponding to electronic diagrams. For each domain, one test contains ideal instances of the symbols while the other three contain increasingly degraded versions of the symbols using the Kanungo's method [6] as in the tests for symbol recognition. We have employed different parameters of the method to provide four levels of degradation: ideal (i.e. without noise), levels 1, 2 and 3. The training tests are composed of 5 drawings each, whereas the final tests are composed of 20. Some examples of these images are shown in Fig. 3.

[1] Systems Evaluation SYnthetic Documents.

[2] `http://mathieu.delalandre.free.fr/projects/sesyd/`

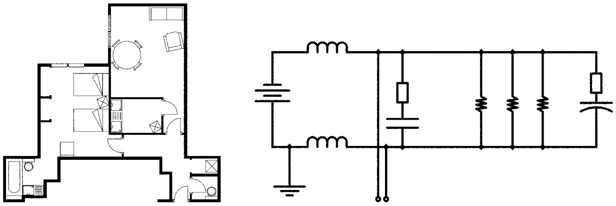

Fig. 3. Examples of images of complete drawings for symbol spotting

3 Evaluation Metrics

For symbol recognition we just used the recognition rate as in previous editions of the contest. For symbol spotting we have adopted the measures proposed in a recent work that redefined classical retrieval performance measures for the case of spotting in graphics recognition [10]. For completeness we recall in the following the definition of these measures as described in the original paper. They are based on the overlapping of the set of polygons describing the ground-truth and the set of polygons returned as a result of spotting. In our case we have constrained the polygons to rectangular bounding boxes of symbols both in the ground-truth and in the results.

Then, being $A(P)$ the area of a set of polygons, being $\bigoplus$ the operator that denotes the intersection of two sets of polygons, P_{rel} the set of polygons in the ground-truth and P_{ret} the set of polygons retrieved by the spotting system, precision P, recall R and F-score F are defined as follows:

$$P = \frac{A(P_{ret} \bigoplus P_{rel})}{A(P_{ret})} \tag{1}$$

$$R = \frac{A(P_{ret} \bigoplus P_{rel})}{A(P_{rel})} \tag{2}$$

$$F = \frac{P \cdot R}{P + R} \tag{3}$$

In addition to these measures two additional measures are defined to evaluate the recognition at symbol level, that is, the percentage of symbols that are found at some degree by the spotting system. This degree of confidence that controls if a symbol has been found is defined in terms of overlapping between the area of the symbol in the ground-truth and the result of the retrieval. Thus, if the overlapping area is above a certain threshold (that is fixed to 75%) of the area of the symbol in the ground-truth the symbol is considered as correctly identified. Then, recognition rate, as the percentage of symbols in the ground-truth correctly identified, and average false positives $AveFP$ as the average number of returned symbols that do not correspond to any ground-truth symbols, are also defined as complementary evaluation metrics.

Table 3. Global results tests on symbol recognition

Test name	Recognition rate
set #1 (50 models)	94,76%
set #2 (100 models)	91,98%
set #3 (150 models)	85,88%
set #4 (cropped images, 36 models)	96,22%

4 Results

There was only one participant in the contest, in both entries, symbol recognition and symbol spotting. The method, based on geometric matching, was developed by Nayef et. al. [9], IUPR research group of the university of Kaiserslautern. The method is based on geometric matching. In the next subsections we will report detailed results for symbol recognition and symbol spotting.

Table 4. Detail of results for set #3 for each kind of deformation

Degradation	Recognition rate
Basic	85,33%
Rotation & scaling	84,84%
Degradation α	88,07%
Degradation β	85,73%
Degradation η	85,67%

4.1 Symbol Recognition

As it has been described in section 2 we generated 4 different tests for symbol recognition. Three of them were created after applying several deformations to an increasing number of symbol models: 50, 100 and 150. The fourth test consisted of images of symbols directly segmented from the drawings including lines of their neighboring elements . The global results for each test are shown in table 3. As expected we can observe that accuracy decreases as the number of symbol models increase. However, the method seems to be robust to segmentation noise. Accuracy for set #4 is higher than in the other tests. These better results could be justified by the lower number of symbol models (only 36) in this test, and also by the fact that the symbols in this test are clean images with no noise at all. The background lines connected to the symbols did not affect the performance since the method works for recognizing symbols in context.

Table 5. Results for set #3 for images with rotation and scaling

Level of degradation	Recognition rate
Rotation	81,07%
Scaling	89,20%
Rotation-Scaling	84,27%

In tables 4–8 we show detailed results for each kind of transformation applied to the images. Table 4 shows the details for each kind of deformation according to the different parameters of Kanungo's degradation model or to the affine transforms (rotation and scaling) applied to the images. We have only included results for set #3 as it is the set with the larger number of symbol models and thus, where differences could be a priori more significant. However, we do no appreciate significant differences in the accuracy for the different kinds of deformations. It is surprising that results for degradation according to parameter α – which generates images with thin lines, as shown in figure 1 – are better than those for the basic set of images with a very slight noise – figure 1 –. A possible explanation could be that the method includes an adaptive preprocessing and noise removal module to deal with different kinds of heavy binary noise.

Table 6. Results for set #3 for different levels of degradation based on parameter α

Level of degradation	Recognition rate
Level 1	88,40%
Level 2	87,73%

Analyzing in more detail the results for each kind of distortion we can observe that the method seems to be more robust to scaling than to rotation – table 5 –. For degradations generated with the Kanungo's model, the performance decreases slightly as the amount of noise increases, although not in a significant way – tables 6–8–.

Table 7. Results for set #3 for different levels of degradation based on parameter β

Level of degradation	Recognition rate
Level 1	85,87%
Level 2	85,60%

4.2 Symbol Spotting

In the spotting tests, participants were asked to spot all instances of all symbol models included in the test. In table 9 we show the results for the tests including images of architectural floorplans with increasing levels of noise. Although there is not a completely linear relation, we can observe a degradation of all the performance indices as the amount of noise increases. This relation is not so clear for images of electrical diagrams 10. At this point, it is probably worth noting

Table 8. Results for set #3 for different levels of degradation based on parameter η

Level of degradation	Recognition rate
Level 1	86,00%
Level 2	85,33%

that the performance of a symbol spotting system can depend on many factors, being the level of noise only one of them. There are other parameters such as the number and location of the symbols that can also have a great influence in the final results. Since all these parameters have been determined at least partially in a random way, we do not have a complete control on the difficulty of every test. In addition, analyzing the results on symbol recognition in the previous section, we can see that the method seems to be quite robust to binary noise.

Table 9. Spotting results for images of architectural floorplans

Test name	Precision	Recall	F-Score	Recognition rate	Average false positives
Set #5 (ideal)	0.62	0.99	0.76	99,31	18,75
Set #6 (level 1)	0.64	0.98	0.77	97,00%	13,68
Set #7 (level 2)	0.62	0.93	0.74	98,80%	13,62
Set #8 (level 3)	0.57	0.98	0.72	97,74%	17,37

Table 10. Spotting results for images of electrical diagrams

Test name	Precision	Recall	F-Score	Recognition rate	Average false positives
Set #9 (ideal)	0.37	0.56	0.45	94,02%	2.66
Set #10 (level 1)	0.44	0.63	0.52	86,27%	3.19
Set #11 (level 2)	0.40	0.61	0.48	85,25%	2.66
Set #12 (level 3)	0.43	0.64	0.51	88,40%	3.76

5 Conclusions

In this paper we have described the framework for the fourth edition of the Symbol Recognition Contest and we have reported the results achieved by the only participant method. This is the first time that the contest includes an entry on symbol spotting.

Concerning symbol recognition, after several editions of the contest we have evolved the dataset including a systematic way of generating several kinds of distortions from a basic set of images of the symbols. We think that this dataset can serve without further significant modifications for future editions of the contest and can become a stable platform for continuous evaluation and comparison of symbol recognition methods, maybe with the only additional inclusion of hand-drawn symbols.

With respect to symbol spotting, this is the first important attempt to provide a complete framework for evaluations, including a significantly large dataset along with a set of performance measures. We feel that the final result is encouraging, although probably some improvement should be done in the creation of the dataset, particularly regarding the generation of noise, to be able to characterize the difficulty of each test. And, obviously, we need to foster participation in the contest to validate the framework.

Acknowledgments. This work has been partially supported by the Spanish projects TIN2009-14633- C03-03, TSI-020400-2011-50 and a CONSOLIDER-INGENIO 2010(CSD2007-00018).

References

1. Aksoy, S., Ye, M., Schauf, M., Song, M., Wang, Y., Haralick, R., Parker, J., Pivovarov, J., Royko, D., Sun, C., Farneback, G.: Algorithm performance contest. In: Proceedings of the 15th International Conference on Pattern Recognition, vol. 4, pp. 870–876 (2000)
2. Chhabra, A.K.: Graphic Symbol Recognition: An Overview. In: Chhabra, A.K., Tombre, K. (eds.) GREC 1997. LNCS, vol. 1389, pp. 68–79. Springer, Heidelberg (1998)
3. Delalandre, M., Valveny, E., Pridmore, T., Karatzas, D.: Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems. Int. J. Doc. Anal. Recognit. 13, 187–207 (2010)
4. Dosch, P., Valveny, E.: Report on the Second Symbol Recognition Contest. In: Liu, W., Lladós, J. (eds.) GREC 2005. LNCS, vol. 3926, pp. 381–397. Springer, Heidelberg (2006)
5. Valveny, E., Dosch, P., Fornes, A., Escalera, S.: Report on the Third Contest on Symbol Recognition. In: Liu, W., Lladós, J., Ogier, J.-M. (eds.) GREC 2007. LNCS, vol. 5046, pp. 321–328. Springer, Heidelberg (2008)
6. Kanungo, T., Haralick, R.M., Stuezle, W., Baird, H.S., Madigan, D.: A statistical, nonparametric methodology for document degradation model validation. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1209–1223 (2000)
7. Lladós, J., Valveny, E., Sánchez, G., Martí, E.: Symbol Recognition: Current Advances and Perspectives. In: Blostein, D., Kwon, Y.-B. (eds.) GREC 2002. LNCS, vol. 2390, pp. 104–128. Springer, Heidelberg (2002)
8. Luqman, M.M., Delalandre, M., Brouard, T., Ramel, J.-Y., Lladós, J.: Fuzzy Intervals for Designing Structural Signature: An Application to Graphic Symbol Recognition. In: Ogier, J.-M., Liu, W., Lladós, J. (eds.) GREC 2009. LNCS, vol. 6020, pp. 12–24. Springer, Heidelberg (2010)
9. Nayef, N., Breuel, T.M.: On the use of geometric matching for both: Isolated symbol recognition and symbol spotting. In: Proceedings of the 9th International Conference on Graphics Recognition, GREC 2011 (2011)
10. Rusiñol, M., Lladós, J.: A performance evaluation protocol for symbol spotting systems in terms of recognition and location indices. International Journal on Document Analysis and Recognition 12, 83–96 (2009)
11. Tombre, K., Tabbone, S., Dosch, P.: Musings on Symbol Recognition. In: Liu, W., Lladós, J. (eds.) GREC 2005. LNCS, vol. 3926, pp. 23–34. Springer, Heidelberg (2006)
12. Valveny, E., Dosch, P., Winstanley, A., Zhou, Y., Yang, S., Yan, L., Wenyin, L., Elliman, D., Delalandre, M., Trupin, E., Adam, S., Ogier, J.M.: A general framework for the evaluation of symbol recognition methods. Int. J. Doc. Anal. Recognit. 9, 59–74 (2007)
13. Valveny, E., Dosch, P.: Symbol Recognition Contest: A Synthesis. In: Lladós, J., Kwon, Y.-B. (eds.) GREC 2003. LNCS, vol. 3088, pp. 368–385. Springer, Heidelberg (2004)