

Performance Evaluation of Symbol Recognition and Spotting Systems: An Overview

Mathieu Delalandre
CVC, Barcelona, Spain
mathieu@cvc.uab.es

Ernest Valveny
CVC,UAB Barcelona, Spain
ernest@cvc.uab.es

Josep Lladós
CVC,UAB Barcelona, Spain
josep@cvc.uab.es

Abstract

This paper deals with the topic of performance evaluation of the symbol recognition & spotting systems. It presents an overview as a result of the work and the discussions undertaken by a working group on this subject. The paper starts by giving a general view of symbol recognition & spotting and performance evaluation. Next, the two main issues of performance evaluation are discussed: groundtruthing and performance characterization. Different problems related to both issues are addressed: groundtruthing of real documents, generation of synthetic documents, degradation models, the use of a priori knowledge, mapping of the groundtruth with the system results, and so on. Open problems arising from this overview are also discussed at the end of the paper.

1 Introduction

Performance evaluation is a particular cross-disciplinary research field in a variety of domains such as Information Retrieval [17], Computer Vision [44], CBIR¹ [35], DIA² [19], etc. Its purpose is to develop full frameworks in order to evaluate, to compare and to select the best-suited methods for a given application. Such a framework includes providing groundtruth and datasets for training and testing, defining a data exchange protocol, defining metrics and providing tools to match the system results with the groundtruth.

This paper deals with the performance evaluation applied to DIA systems. Performance evaluation is a well known topic in DIA since the first works in the early 90's [18]. Performance evaluation frameworks have been defined for several DIA tasks [24], such as table recognition, page segmentation, OCR³, etc. In this paper we are interested in a specific domain of DIA: graphics recognition. Performance

evaluation of graphics recognition systems goes back to the middle of the 90's [27]. The first works focussed on the evaluation of vectorization [37], but in the last years, the interest has moved towards the evaluation of higher-level tasks such as symbol recognition and spotting [46], especially with the organization of three International Contests on Symbol Recognition [1] [49] [15].

This paper reports a summary of the work and the discussions undertaken by a working group about performance evaluation of symbol recognition/spotting. The purpose of this working group was to review past works on this topic, but also to propose a kind of "to do list" for future research. Then, this paper is a combination of overview and guidelines for research.

Symbol recognition is an active topic in the field of graphics document understanding. Several surveys [8] [11] [31] [47] review existing works on logical diagrams, engineering drawings, maps, etc. In a very general way [31], a symbol can be defined as "*a graphical entity with a particular meaning in the context of an specific application domain*" and then symbol recognition as "*a particular application of the general problem of pattern recognition, in which an unknown input pattern (i.e. input image) is classified as belonging to one of the relevant classes (i.e. pre-defined symbols) in the application domain*". So, as any pattern recognition application, symbol recognition relies on two types of input data (Figure 1): test documents and learning data. Then, the system has to localize and to recognize the symbols in the document.

One of the major problems of symbol recognition is to combine segmentation and recognition. This problem is known as the segmentation/recognition paradigm in the literature [57]: a system should segment before recognizing but, at the same time, some kind of recognition may be necessary for the segmentation. In order to overcome this paradox, research has been directed to symbol spotting [46]. Since research on symbol spotting is just starting, it is still a little ambiguous to define "what a spotting method is". In [47] it is defined as "*a way to efficiently localize possible symbols and limit the computational complexity, without us-*

¹Content Based Image Retrieval

²Document Image Analysis

³Optical Character Recognition

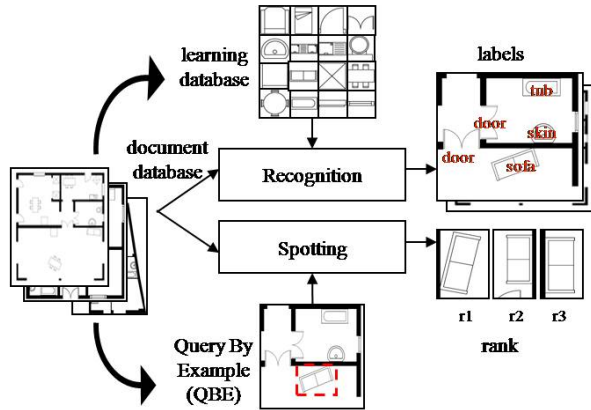


Figure 1. Recognition/spotting of symbols

ing full recognition methods”. So, spotting is presented as a kind of middle-line technique combining recognition and segmentation. Symbol spotting systems can also be viewed as CBIR systems. Indeed, most of the existing symbol spotting systems [14] [43] [61] [32] [40] [41] work in a way similar to CBIR (Figure 1). Spotting is initiated with a query selected from a drawing by the user, what we call a QBE⁴. Then, this QBE is used as a model to find similar symbols in the document database. At the end, the system provides a ranked list of similar symbols along with their localization data (i.e. url of the source document with the coordinates of the symbol).

In both cases (spotting and recognition), a hard problem is how to obtain and compare experimental results from existing systems. Traditionally, this step was done independently for every system [8] [31] [47], by comparing manually the results with the original images and checking the recognition errors. This process was unreliable as it raises conflicts of interest and does not provide relevant results. Moreover, it does not allow to compare different systems and test them with large amounts of data. In order to solve these problems research has been initiated over the last few years on the performance evaluation of symbol recognition/spotting systems [1] [49] [15] [48].

Due to the heterogeneity of fields related to performance evaluation [44] [35] [19] there is not a common definition of “what a performance evaluation framework is”. However, two main tasks are usually identified (Figure 2): groundtruthing, which provides the reference data to be used in the evaluation, and performance characterization, which determines how to match the results of the system with the groundtruth to give a measure of the performance of the system. In the follow-up we analyze these two issues in sections 2 and 3. We will conclude in the section 4 to discussing some open problems arising from this overview.

⁴Query by Example

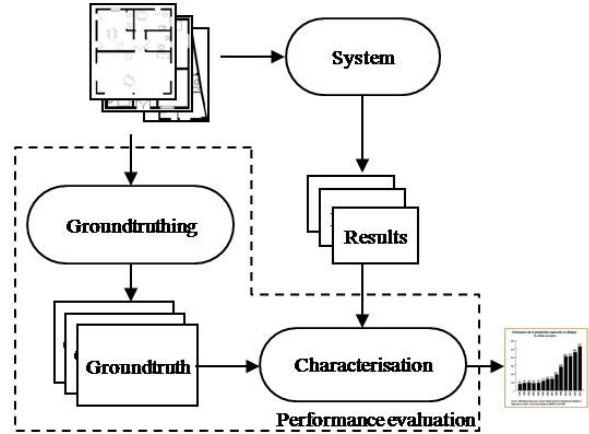


Figure 2. Performance evaluation

2 Groundtruthing

2.1 Introduction

The first step to evaluate any graphics recognition application is to provide test documents with their corresponding groundtruth [33]. In the last years several groundtruthing systems have been proposed: [15], [54], [1], [60], [48] and [13]. We will present and discuss all of them in the next subsections 2.2 and 2.3. In addition to that, to support this discussion we compare these systems in the Table 1 according to different criteria: quickness of the groundtruthing process, realism of test documents, reliability (which level of error in the groundtruth), number of symbol per image, connected or segmented symbols, possibility to add noise.

	Quickness	Realism	Reliability	Number	Connected	Noise-effect	
-- weak ++ good							
[Dosch'06 [15]	--	++	-	many	yes	no	bottom-up
[Yan'04 [54]	-	++	--	many	yes	no	
[Chhabra'98 [9]	+	++	+	many	yes	no	top-down
[Aksoy'00 [1]	++	--	++	many	no	yes	
[Zhai'03 [60]	++	--	++	one	no	yes	
[Valveny'07 [48]	++	--	++	one	no	yes	
[Delalandre'07 [13]	++	+	++	many	yes	no	

Table 1. Comparison of groundtruthing methods

2.2 The bottom-up approach

A natural approach is to define the groundtruth from scanned paper documents. Then, a GUI⁵ can be used by human operators in order to edit manually the groundtruth. Thus, the groundtruthing starts from low-level data (e.g. raster images or sets of unstructured vectors) in order to provide high level descriptions of the content (e.g. graphical labels, region of interest, etc.). We will refer here to this approach as *bottom-up*. As the groundtruth is edited by humans, it is necessary to do this task collaboratively with different operators [33]. In this way errors produced by a single operator can be avoided.

In the past, this approach has been mainly applied to the evaluation of layout analysis and OCR [56] [30] [4]. Concerning graphical documents, only the EPEIRES⁶ platform exists up to day [15]. It is presented in the Figure 3. This system is based on a collaborative approach using two main components: a GUI to edit the groundtruth connected to an information system. The operators obtain from the system the images to annotate and the associated symbol models. The groundtruthing is performed by mapping (moving, rotating and scaling) transparent bounded models on the document using the GUI. The information system allows to collaboratively validate the groundtruth. Experts check the groundtruth generated by the operator by emitting alerts in the case of errors.

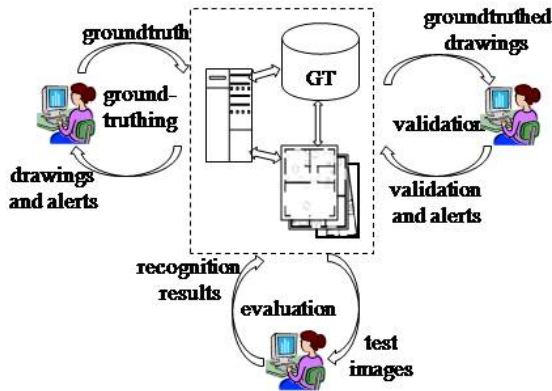


Figure 3. The EPEIRES system

Despite this existing platform a problem still remains: the time and cost required to edit the groundtruth. Existing works [56] [33] [30] [4] highlight that, in most of the cases, the groundtruthing effort makes very hard the creation of large databases. An alternative approach to avoid this problem is semi-automatic groundtruthing. In this case, the key idea is to use a recognition method to obtain an initial version of the groundtruth. Then, the user has only to validate

⁵Graphics User Interface

⁶<http://epeires.loria.fr/>

and correct the recognition results in order to provide the final groundtruth. This approach has already been used in other applications like OCR [28], layout analysis [45], chart recognition [55], etc. Concerning symbol recognition only the system described in [54] has been proposed until now. This system recognizes engineering drawings using a case-based approach. The user starts by targeting a graphical object (i.e. a symbol) in an engineering drawing. Then, the system learns a graphical model of this object and uses it to localize and recognize similar objects in the drawing. The Figure 4 describes this model. During the learning process, the system also takes into account user feedback on positive and negative examples. It modifies the original knowledge by computing tolerances about the primitives and their relations (length, angle, line number, etc.).

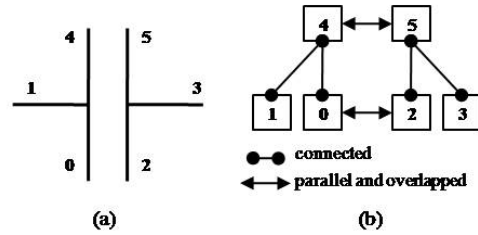


Figure 4. System of [54]

(a) symbol (b) line graph

2.3 The top-down approach

The bottom-up approach results in realistic data but raises complex problems: how to define the ground-truth, how to deal with the errors introduced by users, the delay and the cost of the groundtruth acquisition, etc. In many cases these problems render the approach impractical to constitute large databases. Other works [9] [1] [60] [48] [13] are based on a different approach. The key idea is to use documents of high-level content (like the vector graphics “DXF, SVG, CGM, etc.”) and to convert them into images. In this way, they can take advantage of a groundtruth already existing: it is not necessary to re-define it. We will refer to this kind of approach as *top-down*. The systems using a top-down approach can be distinguished in two categories in the literature: using CAD⁷ or synthetic data.

In the first kind of systems, the groundtruthing process works with real-life documents edited with CAD tools (like AutoCad). These documents are then converted into images to create evaluation tests. This method has been used until now to evaluate the processes of raster to vector conversion [9]. However, it could be easily extended to symbol recognition by using the symbol layer of the CAD files. The main

⁷Computer Aided Design

difficulty of this approach is to collect the initial CAD documents [38]. This process must deal with different aspects: to collect the documents and their copyrights, to record the documents (to define single id, to find duplicates), to valid the format for the storage and to convert it to a standard format when necessary, to edit metadata about the documents in order to organize the database, etc.

A complementary approach, which avoids these difficulties, is to create and to use synthetic documents. Here, the test documents are built by an automatic system which combines pre-defined models of document components in a pseudo-random way. Test documents and ground-truth can therefore be produced simultaneously. In addition, a large number of documents can be generated easily and with limited user involvement. Several systems have been proposed in the literature [1], [60], [48] and [13]. Figure 5 gives some examples of documents produced by these systems.

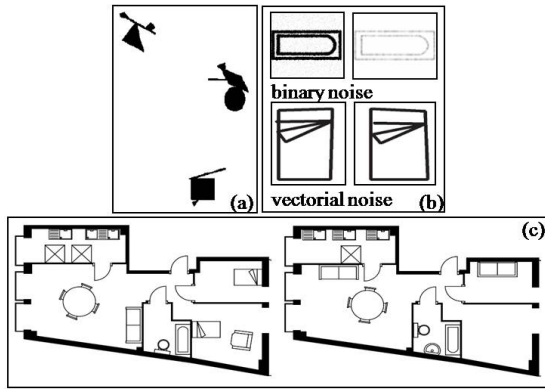


Figure 5. Examples of synthetic document

(a) random symbol set (b) segmented symbol
(c) document instances

The system described in [1] employs an approach to build documents composed of multiple unconnected symbols. Figure 5 (a) gives an example of such a document. Each symbol is composed of primitives (circles, lines, squares, etc.) randomly selected and mildly overlapped. Next, they are placed on the image at a random location and without overlapping with the bounding boxes of other symbols. The systems proposed by [60] and [48] support the generation of degraded images of segmented symbols as shown in the Figure 5 (b). In these systems, the models of the symbols are described in a vector graphics format. The vector graphics files are then converted into images. The system uses a random selection process to select a model from the model database, and apply to it a set of scaling and rotation operations. The authors in [13] extend the systems of [60] [48] for the generation of whole documents (drawings, maps, diagrams, etc.). They exploit the layer property of graphical documents in order to po-

sition sets of symbols in different ways on the same background. They obtain document instances as those shown in the Figure 5 (c). The positioning of the symbols is based on the use of some constraints that define how a control point on the model matches a positioning one defined on the background. In order to allow a flexible positioning each constraint also permits previous scaling and rotation of the symbol. The control point can be defined anywhere inside the symbol by using a couple of polar coordinates. At last, the positioning point can be randomly generated on regions (lines or polygons) previously defined on the background.

In all these systems an important issue is the generation of images resembling as much as possible to real documents. In this sense, as real documents are usually degraded due to multiple sources of noise, it is necessary to use degradation models in the process of generation that permits to simulate this noise. We can distinguish two main sources of degradation: the degradation due to the printing and/or acquisition of the documents, and the degradation due to the process of generation of the documents (in this case, the main source of variability is handwriting).

For the first kind of degradation, two different models of degradation have been proposed [26] [5] that try to reproduce the process of printing and acquisition. Both models have been used in the generation of synthetic data in different applications of DIA, specially OCR. In all the past contests on symbol recognition [1] [49] [15], the method proposed in [26] has been used to generate degraded images (some examples can be seen in the Figure 5 (b)).

For the second kind of degradation few work exists. In the two past editions of the contest on symbol recognition [49] [15] the shape of the symbols has been distorted using the method proposed by [50]. It employs a probabilistic model that modifies the location, the orientation and the length of the lines of the symbol. This probabilistic model is learned from a training set composed of real images based on the Active Shape Models [10]. Some examples of the images that are generated using this model can be seen in the Figure 5 (b).

2.4 *A priori* knowledge

Before processing any input document image, an automatic processing system needs some *a priori* knowledge. The *a priori* knowledge depends on the application (segmentation, recognition, OCR, etc.). In the case of recognition it corresponds to learning databases for training. For spotting, a set of QBE. So, the performance evaluation framework has also to provide the dataset corresponding to this *a priori* knowledge.

Concerning recognition, these *a priori* knowledge depend of the used method. The methods can be classified according to two main families [8] [11] [31]: statistic and

syntactic & structural. In the second case, methods usually work at the graph level where it is difficult to do a learning step [22]. In order to take into account this specificity, past Contests [1] [49] [15] provided two kinds of training data: the usual learning databases and also sets of ideal models (i.e. the ideal shapes without noise). These ideal models permitted to provide a representative symbol per class for methods that do not need the learning step. These Contests were applied to the recognition of segmented symbol images. In the case of segmentation and recognition of symbols in whole documents the learning step is slightly different because the systems have to learn about the context where the symbols can appear. This context corresponds to the other graphical elements surrounding the symbol in the document (e.g. background, neighboring symbols, text, etc.). The Figure 6 shows some examples of symbols in the context of whole documents. These contextual information could be used during the learning to make the recognition more robust [54], or to develop segmentation strategies [20] [59]. For these reasons, the recognition from complete documents involves training using images of complete documents with several instances per class. This will permit the systems to apply a rejection strategy in order to improve their recognition abilities. These points are important issues in machine learning [58] that have not been considered up to day in the past symbol recognition Contests.

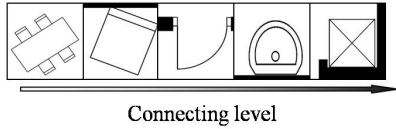


Figure 6. symbols in context

Spotting is a different case because it does not raise on training data but on sets of QBE to initiate the retrieval. The existing works on spotting [14] [43] [61] [32] [40] [41] show experimental results based on some QBE defined by the authors themselves. However, these QBE could have a very different precision which can have a large influence in the spotting results. There is no common idea of what a mean QBE is. However the existing works [14] [43] [61] [32] [40] [41] argue that the users do crops as illustrated in the Figure 7 (a). It is then important to have a previous idea about the precision, and to test the systems on a large number of QBE to make the evaluation more accurate. This raises the problem of collecting an initial set of QBE: doing it manually would take more time than generating the groundtruth of symbols (more than one QBE could be produced from the same symbol location). So, methods for automatic generation of QBE have to be used. Some previous works [50] applied to the hand-sketched vectorial distortion could be a way to approach this problem.

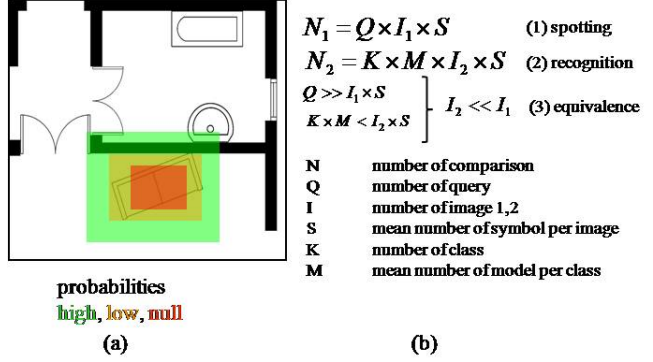


Figure 7. Spotting and QBE

(a) cropping (b) recognition vs spotting complexity

Another problem related to spotting is complexity. Spotting is mainly different from recognition because a query has to access the full database of documents while respecting a real time constraint. Thus, the number of comparisons to perform with the test databases is more important. The Figure 7 (b) gives some considerations about this problem by comparing the complexity of spotting and recognition. Because the complexity is an integral aspect of spotting the systems generally use some kind of heuristics (hash table [41], node seeds [40], dendrogram [61], etc.). However, from the point of view of performance evaluation it is important to keep care of the amount of data to be processed by the systems. In the case of spotting a trade-off should be considered between the number of test images and the number of QBE.

3 Performance Characterization

Once the groundtruthing is done it is possible to test the systems. The final evaluation is achieved by comparing the system results with the groundtruth using a performance characterization method. The objective is to detect good and bad recognition/spotting cases in order to compute performance measures about the systems. As these systems rely on a classification process, it is possible to take benefit of the performance evaluation works done in this field [29]. Usual evaluation tools are the recognition rate, the cross validation, the confusion matrix, precision & recall, the ROC⁸ curve, the F-measure, etc. They are well-known in the domains of Pattern Recognition [16] and Information Retrieval [17] and used in various fields like Computer Vision [44], CBIR [35], DIA [7], etc.

Concerning symbol recognition a contribution using such tools has been proposed recently by [51]. The authors apply the measures of homogeneity, separability, recogni-

⁸Receiver Operating Characteristic

tion rate and precision-recall to evaluate a collection of shape descriptors applied to the recognition of segmented symbol images. However, with whole documents this task becomes harder. Indeed, the comparison of the groundtruth with the results cannot be done between segmented symbols, but between sets of symbols. These sets could be of different size, and large gaps could appear between the locations of symbols. So, before doing any characterization it is necessary to find the correspondences between the system results and the groundtruth as illustrated in the Figure 8. We will refer to this process as *mapping*.

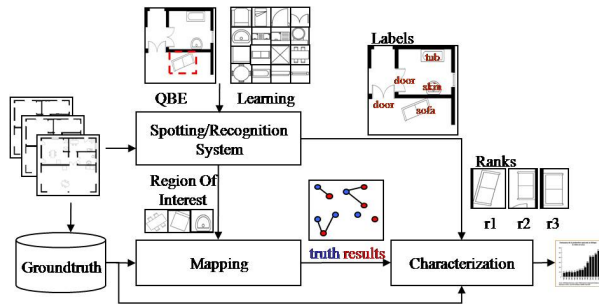


Figure 8. Performance characterization

In the field of graphics recognition past work has been proposed on mapping to evaluate the processes of raster to vector conversion. In [37] five mapping cases have been defined between the groundtruthed and result vectors (Figure 9). Algorithms supporting these cases to measure the accuracy of vectorization systems have been proposed by [53] and [37]. However, mapping of symbols is different because it does not aim to match a distance between vector sets, but to determine the overlapping cases between the detected symbols and the groundtruth. These overlapping cases must be detected by comparing the locations of all the symbols at the same time. The final evaluation results will be obtained by computing the rates of correct recognition and spotting during the characterization.

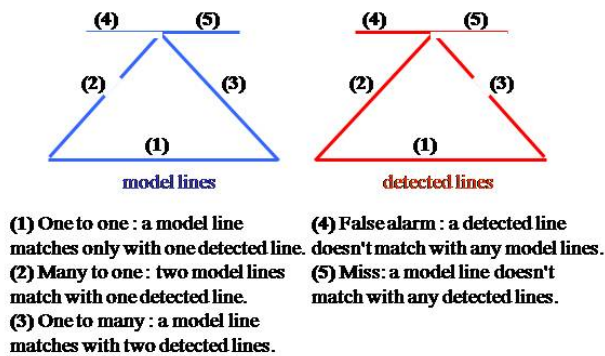


Figure 9. Mapping cases of [37]

To the best of our knowledge, it does not exist any work on mapping for the evaluation of symbol recognition & spotting. However, several related works have been proposed in other fields like in OCR [25], layout analysis [2], text/graphics segmentation [52], handwriting segmentation [42], etc. In these works, the major question is to determine how to describe the areas to be mapped (both in the groundtruth and in the results). Different possibilities have been explored (Figure 10): using wrappers [52], contours [2] and label maps [6]. As the performance of an algorithm depends to the objects to handle, each description is a trade-off between the accuracy and the complexity of the characterization.

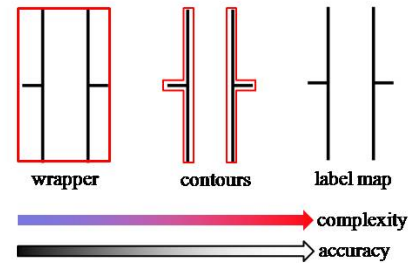


Figure 10. Description of areas

A first way to perform mapping is to use wrappers. A famous example of such a wrapper is the bounding box; others are the ellipsis, the parallelogram, etc. The overlapping rates are obtained using well-known mathematical functions because the wrappers are common geometrical shapes. Several systems have used this approach in the past. In [52] orientated bounding boxes are used to match characters for the evaluation of text/graphics separation algorithms. In layout analysis [12] [34] rectangular blocks are used to describe the page components (paragraph, title, etc.). The groundtruth is matched with the layout analysis results to detect over and under segmentation cases. For the evaluation of OCR, the systems of [21] and [25] map together sets of character bounding boxes by applying geometric transformations.

The drawback of using wrappers is the precision. It will depend a lot on the shape of the symbol as illustrated in the Figure 11. In order to make mapping more precise another solution is to use contours. To the best of our knowledge, only the system of [2] has been proposed to work at such a level. This system has been used during the fourth international Page Segmentation Competition [3]. The major problem of using contours is the complexity as the comparison of polygons has a polynomial time. In order to avoid this problem, the system [2] uses isothetic polygons as shown in the Figure 12. Thus, polygons are compared using their intervals. These intervals are defined as maximal rectangles that can be fitted horizontally inside a region (starting at a

given point on a vertical edge), spanning the whole width of the region. Figure 12 represents the obtained intervals between two segmented regions and a groundtruthed one.

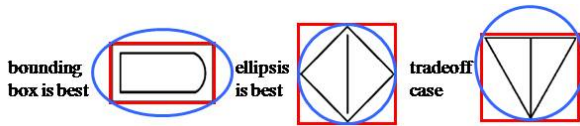


Figure 11. Wrapper sensitivity to the models

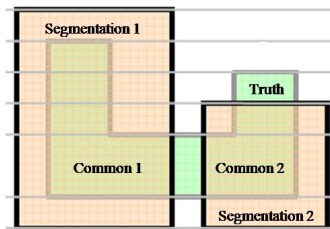


Figure 12. Mapping of [2]

A last way to do mapping is to employ label maps. In these label maps, each label represents a specific zone. Such an approach has been used for handwriting segmentation [6] [36], layout analysis [42] and document image retrieval [23]. The comparison of the groundtruth and the results can simply be done by finding the number of common pixel between the groundtruth and the result label maps [36] [23]. A more complex metric is proposed by [6] and used in the system of [42]. In this case, the groundtruth and the result label maps are represented with a weighted bipartite graph called pixel-correspondence graph. In this graph, the nodes represent the segmented regions (i.e. a groundtruth character or a character hypothesis) and the edges the overlapping rates (when the overlapping exists). A perfect segmentation case will correspond to a bipartite equivalence in the graph (i.e. same number of nodes and every node on either side of the graph has exactly one edge).

4 Discussion

In this paper we have presented an overview about the performance evaluation of symbol recognition & spotting systems. Main conclusions and open issues arising from this overview are discussed in this section.

In the last years, some works have been undertaken to provide groundtruthed databases in order to evaluate symbol recognition & spotting methods, using real as well as synthetic data [1] [60] [49] [15] [48] [13]. These works have been applied first to segmented symbols [1] [60] [49] [48] and recently extended to connected symbols in whole

documents [15] [13], which is the original goal of the groundtruthing systems. Despite this progress, several open issues still remain. On the one hand, the time needed to collect and groundtruth real-life documents makes complex their use in most of the cases. On the other hand, synthetic methods have difficulties to reproduce the variability of real documents. Thus, further works have to be done in order to speed-up the groundtruthing process and to make the synthetic data more real. As these two approaches have intrinsic drawbacks and advantages, they should be combined in the future evaluation campaigns.

Another open question is to address the machine learning issues in the performance evaluation of symbol recognition & spotting systems. The size and the “quality” of the training data have a great impact on the system results. In the field of machine learning this is a very important issue that has not been considered at large up to day in the past Contests of symbol recognition [1] [49] [15]. The participants should mention explicitly what are the training datasets they employ, and should provide experiments about their systems using different sets. In the same manner pre-processing chains are also an important feature to take care of. So, the participants should describe their method precisely (which algorithms for preprocessing have been used and which *a priori* knowledge has been taken into account). A previous methodology has been proposed to describe graphics recognition systems at a knowledge level [39]. It could be a way to approach this problem.

A last problem concerns the characterization (i.e. the final evaluation of systems). It has been done in the past Contests [49] [15] using results obtained from segmented symbol images. However, up to day no contribution have been proposed with the whole documents. It should be one challenge for the graphics recognition community to propose such methods in a near future. The major problem of this step is mapping of the groundtruth with the system results. Past related works can be found in the performance evaluation of layout analysis and OCR methods [52] [25] [2]. The graphics recognition community should take benefit of these contributions to initiate works on this topic.

5 Acknowledgement

The authors wish to thank the members of the working group for our exchanges and discussions on this topic: Alicia Fornes, Dimosthenis Karatzas, Hervé Locteau, Jean-Pierre Salmon, Jean-Yves Ramel, Marçal Rusinol, Philippe Dosch, Rashid Qureshi and Tony Pridmore. Work partially supported by the Spanish project TIN2006-15694-C02-02, and by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

References

- [1] S. Aksoy and al. Algorithm performance contest. In *International Conference on Pattern Recognition (ICPR)*, volume 4, pages 870–876, 2000.
- [2] A. Antonacopoulos and D. Bridson. Performance analysis framework for layout analysis methods. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 1258–1262, 2007.
- [3] A. Antonacopoulos, B. Gatos, and D. Bridson. Icdar2007 page segmentation competition. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 1279–1283, 2007.
- [4] A. Antonacopoulos, D. Karatzas, and D. Bridson. Ground truth for layout analysis performance evaluation. In *Workshop on Document Analysis (DAS)*, volume 3872 of *Lecture Notes in Computer Science (LNCS)*, pages 302–311, 2006.
- [5] H. Baird. Document image defect models and their uses. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 62–67, 1993.
- [6] T. Breuel. Representations and metrics for off-line handwriting segmentation. In *International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, pages 428–433, 2002.
- [7] N. Chen and D. Blostein. A survey of document image classification: problem statement, classifier architecture and performance evaluation. *International Journal on Document Analysis and Recognition (IJ DAR)*, 10(1):1–16, 2007.
- [8] A. Chhabra. Graphic symbol recognition: An overview. In *Workshop on Graphics Recognition (GREC)*, volume 1389 of *Lecture Notes in Computer Science (LNCS)*, pages 68–79, 1998.
- [9] A. Chhabra and I. Phillips. The second international graphics recognition contest - raster to vector conversion : A report. In *Workshop on Graphics Recognition (GREC)*, volume 1389 of *Lecture Notes in Computer Science (LNCS)*, pages 390–410, 1998.
- [10] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding (CVIU)*, 61(1):38–59, 1995.
- [11] L. Cordella and M. Vento. Symbol and shape recognition. In *Workshop on Graphics Recognition (GREC)*, volume 1941 of *Lecture Notes in Computer Science (LNCS)*, pages 167–182, 1999.
- [12] A. Das, S. Saha, and B. Chanda. An empirical measure of the performance of a document image segmentation algorithm. *International Journal of Document Analysis and Recognition (IJ DAR)*, 4:183–190, 2002.
- [13] M. Delalandre, T. Pridmore, E. Valveny, E. Trupin, and H. Locteau. Building synthetic graphical documents for performance evaluation. In *Workshop on Graphics Recognition (GREC)*, pages 84–87, 2007.
- [14] P. Dosch and J. Lladós. Vectorial signatures for symbol discrimination. In *Workshop on Graphics Recognition (GREC)*, volume 3088 of *Lecture Notes in Computer Science (LNCS)*, pages 154–165, 2004.
- [15] P. Dosch and E. Valveny. Report on the second symbol recognition contest. In *Workshop on Graphics Recognition (GREC)*, volume 3926 of *Lecture Notes in Computer Science (LNCS)*, pages 381–397, 2006.
- [16] M. Friedman and A. Kandel. *Introduction to Pattern Recognition : Statistical, Structural, Neural and Fuzzy Logic Approaches*. Number ISBN-10: 9810233124. World Scientific Publishing Company, 1999.
- [17] E. Greengrass. Information retrieval: A survey. Technical Report TR-R52-008-001, Center for Architectures for Data-Driven Information Processing (CADIP), University of Maryland, US, 2000.
- [18] R. Haralick. Performance characterization in image analysis: Thinning, a case in point. *Pattern Recognition Letters (PRL)*, 13(1):5–12, 1992.
- [19] R. Haralick. Performance evaluation of document image algorithms. In *Workshop on Graphics Recognition (GREC)*, volume 1941 of *Lecture Notes in Computer Science (LNCS)*, pages 315–323, 2000.
- [20] J. D. Hartog. Knowledge based interpretation of utility maps. *Computer Vision and Image Understanding (CVIU)*, 63(1):105–117, 1996.
- [21] J. Hobby. Matching document images with ground truth. *International Journal on Document Analysis and Recognition (IJ DAR)*, 1(1):52–61, 1998.
- [22] X. Jiang, A. Munger, and H. Bunke. On median graphs : Properties, algorithms, and applications. *Pattern Analysis and Machine Intelligence (PAMI)*, 23(10):1144–1151, 2001.
- [23] N. Journet, J. Ramel, R. Mullot, and V. Eglin. A proposition of retrieval tools for historical document images libraries. In *International Conference on Document Analysis and Recognition (ICDAR)*, volume 2, pages 1053–1057, 2007.
- [24] T. Kanungo, H. Baird, and R. Haralick. *International Journal on Document Analysis and Recognition (IJ DAR)*, Special Issue on "Performance Evaluation: Theory, Practice, and Impact", volume 4/3. Springer, 2002.
- [25] T. Kanungo and R. Haralick. An automatic closed-loop methodology for generating character groundtruth for scanned documents. *Pattern Analysis and Machine Intelligence (PAMI)*, 21(2):179–183, 1999.
- [26] T. Kanungo, R. Haralick, h.S. Baird, W. Stuehle, and D. M. and. A statistical, nonparametric methodology for document degradation model validation, 2000.
- [27] R. Kasturi and I. Phillips. The first international graphics recognition contest-dashed-line recognition competition. In *Workshop on Graphics Recognition (GREC)*, volume 1072 of *Lecture Notes in Computer Science (LNCS)*, 1996.
- [28] D. W. Kim and T. Kanungo. A point matching algorithm for automatic generation of groundtruth for document images. In *Workshop on Document Analysis Systems (DAS)*, pages 475–485, 2000.
- [29] N. Lavesson. Evaluation of classifier performance and the impact of learning algorithm parameters. Master's thesis, Department of Software Engineering and Computer Science, Blekinge Institute of Technology, Sweden, 2003.
- [30] C. Lee and T. Kanungo. The architecture of trueviz: A groundtruth/metadata editing and visualizing toolkit. *Pattern Recognition (PR)*, 36(3):811–825, 2003.
- [31] J. Lladós, E. Valveny, G. Sánchez, and E. Martí. Symbol recognition : Current advances and perspectives. In *Workshop on Graphics Recognition (GREC)*, volume 2390 of *Lecture Notes in Computer Science (LNCS)*, pages 104–127, 2002.

- [32] H. Locteau, S. Adam, E. Trupin, J. Labiche, and P. Heroux. Symbol spotting using full visibility graph representation. In *Workshop on Graphics Recognition (GREC)*, pages 49–50, 2007.
- [33] D. Lopresti and G. Nagy. Issues in ground-truthing graphic documents. In *Workshop on Graphics Recognition (GREC)*, volume 2390 of *Lecture Notes in Computer Science (LNCS)*, pages 46–66, 2002.
- [34] S. Mao and T. Kanungo. Software architecture of pset: a page segmentation evaluation toolkit. *International Journal on Document Analysis and Recognition (IJ DAR)*, 4:205–217, 2002.
- [35] H. Muller, W. Muller, D. Squire, S. Marchand-Maillet, and T. Pun. Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters (PRL)*, 22(5):593–601, 2001.
- [36] S. Nicolas, T. Paquet, and L. Heutte. Complex handwritten page segmentation using contextual models. In *Document Image Analysis for Libraries (DIAL)*, pages 46–57, 2006.
- [37] I. Phillips and A. Chhabra. Empirical performance evaluation of graphics recognition systems. *Pattern Analysis and Machine Intelligence (PAMI)*, 21(9):849–870, 1999.
- [38] I. Phillips, J. Ha, R. Haralick, and D. Dori. The implementation methodology for the cd-rom english document database. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 484–487, 1993.
- [39] T. Pridmore, A. Darwish, and D. Elliman. Interpreting line drawing images: A knowledge level perspective. In *Workshop on Graphics Recognition (GREC)*, volume 2390 of *Lecture Notes in Computer Science (LNCS)*, pages 245–255, 2002.
- [40] R. Qureshi, J. Ramel, D. Barret, and H. Cardot. Symbol spotting in graphical documents using graph representations. In *Workshop on Graphics Recognition (GREC)*, pages 39–40, 2007.
- [41] M. Rusinol and J. Lladós. A region-based hashing approach for symbol spotting in technical documents. In *Workshop on Graphics Recognition (GREC)*, 2007.
- [42] F. Shafait, D. Keysers, and T. Breuel. Pixel-accurate representation and evaluation of page segmentation in document images. In *International Conference on Pattern Recognition (ICPR)*, pages 872–875, 2006.
- [43] S. Tabbone, L. Wendling, and D. Zuwala. A hybrid approach to detect graphical symbols in documents. In *Workshop on Document Analysis Systems (DAS)*, volume 3163 of *Lecture Notes in Computer Science (LNCS)*, pages 342–353, 2004.
- [44] N. A. Thacker, A. F. Clark, J. Barron, R. Beveridge, C. Clark, P. Courtney, W. Crum, and V. Ramesh. Performance characterisation in computer vision: A guide to best practices. Technical Report 2005-009, Medical School, University of Manchester, Manchester, UK, 2005.
- [45] G. F. G. Thoma. Ground truth data for document image analysis. In *Symposium on Document Image Understanding and Technology (SDIUT)*, pages 199–205, 2003.
- [46] K. Tombre and B. Lamiroy. Graphics recognition - from re-engineering to retrieval. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 148–155, 2003.
- [47] K. Tombre, S. Tabbone, and P. Dosch. Musings on symbol recognition. In *Workshop on Graphics Recognition (GREC)*, volume 3926 of *Lecture Notes in Computer Science (LNCS)*, pages 23–34, 2005.
- [48] E. Valveny and al. A general framework for the evaluation of symbol recognition methods. *International Journal on Document Analysis and Recognition (IJ DAR)*, 1(9):59–74, 2007.
- [49] E. Valveny and P. Dosch. Symbol recognition contest: A synthesis. In *Workshop on Graphics Recognition (GREC)*, volume 3088 of *Lecture Notes in Computer Science (LNCS)*, pages 368–386, 2004.
- [50] E. Valveny and E. Martí. A model for image generation and symbol recognition through the deformation of lineal shapes. *Pattern Recognition Letters (PRL)*, 24(15):2857–2867, 2003.
- [51] E. Valveny, S. Tabbone, O. Ramos, and E. Philippot. Performance characterization of shape descriptors for symbol representation. In *Workshop on Graphics Recognition (GREC)*, 2007.
- [52] L. Wenyn and D. Dori. A proposed scheme for performance evaluation of graphics/text separation algorithms. In *Workshop on Graphics Recognition (GREC)*, volume 1389 of *Lecture Notes in Computer Science (LNCS)*, pages 335–346, 1997.
- [53] L. Wenyn and D. Dori. A protocol for performance evaluation of line detection algorithms. *Machine Vision and Applications*, 9:240–250, 1997.
- [54] L. Yan and L. Wenyn. Interactive recognizing graphic objects in engineering drawings. In *Workshop on Graphics Recognition (GREC)*, volume 3088 of *Lecture Notes in Computer Science (LNCS)*, pages 126–137, 2004.
- [55] L. Yang, W. Huang, and C. Tan. Semi-automatic ground truth generation for chart image recognition. In *Workshop on Document Analysis Systems (DAS)*, volume 3872 of *Lecture Notes in Computer Science (LNCS)*, pages 324–335, 2006.
- [56] B. Yanikoglus and L. Vincent. Pink panther: a complete environment for ground-truthing and benchmarking document page segmentation. *Pattern Recognition (PR)*, 31(9):1191–1204, 1998.
- [57] S. Yoon, G. Kim, Y. Choi, and Y. Lee. New paradigm for segmentation and recognition. In *Workshop on Graphics Recognition (GREC)*, pages 216–225, 2001.
- [58] L. Younes. Introduction to machine learning. Technical report, Department of Applied Mathematics and Statistics, Center for Imaging Science, Johns Hopkins University, Baltimore, USA, 2008.
- [59] Y. Yu, A. Samal., and S. Seth. A system for recognizing a large class of engineering drawing. *Pattern Analysis and Machine Intelligence (PAMI)*, 19(8):868–890, 1997.
- [60] J. Zhai, L. Wenyn, D. Dori, and Q. Li. A line drawings degradation model for performance characterization. In *International Conference on Document Analysis And Recognition (ICDAR)*, pages 1020–1024, 2003.
- [61] D. Zuwala and S. Tabbone. A method for symbol spotting in graphical documents. In *Workshop on Document Analysis Systems (DAS)*, volume 3872 of *Lecture Notes in Computer Science (LNCS)*, pages 518–528, 2006.