

# A Performance Characterization Algorithm for Symbol Localization

Mathieu Delalandre<sup>1</sup>, Jean-Yves Ramel<sup>2</sup>, Ernest Valveny<sup>1</sup> and Muhammad Muzzamil Luqman<sup>2</sup>

<sup>1</sup> *Computer Vision Center (CVC), 08193 Bellaterra (Barcelona), Spain*  
*mathieu@cvc.uab.es; ernest@cvc.uab.es*

<sup>2</sup> *Laboratoire d'Informatique (LI), 37200 Tours, France*  
*jean-yves.ramel@univ-tours.fr; muhammadmuzzamil.luqman@etu.univ-tours.fr*

## Abstract

In this paper we present an algorithm for performance characterization of symbol localization systems. This algorithm aims to be more “generic” and “fuzzy” to characterize the performance. It exploits only single points as the results of localization and compare them with the groundtruth, using information about context. Probability scores are computed for each localization point, depending on the spatial distribution of the regions in the groundtruth. Final characterization results are given with a detection rates/probability error plot, describing the sets of possible interpretations of the localization results. We present experiments and results done with the symbol localization system of [1], using a synthetic dataset of floorplans (100 images, 2500 symbols). We conclude about the performance of this system, in terms of localization accuracy and precision level (false alarms and multiple detections).

*Keywords:* symbol localization, groundtruth, performance characterization, context information

## 1 Introduction

In recent years there has been a noticeable shift of attention, within the graphics recognition community, towards performance evaluation of symbol recognition systems. This interest has led to the organization of several international contests and development of performance evaluation frameworks [2]. However, to date, this work has been focussed on recognition of isolated symbols. It didn't take into account the localization of symbols in real documents. Indeed, symbol localization constitutes a hard research gap, both for recognition and performance evaluation tasks.

Different research works have been recently undertaken to fill this gap [3]. Groundtruthing frameworks for complete documents and datasets have been proposed in [4, 2], and different effective systems working at localization level in [5, 1, 4]. The key problem is now to define characterization algorithms working in a localization context. Indeed the characterization of localization in complete documents is harder, as comparison of results with groundtruth needs to be done between sets of symbols. These sets could be of different size, and significant differences could appear between the localizations of a result symbol (provided by a given system) and the corresponding one in the groundtruth. Characterization metrics must then be reformulated to take these specificities into account.

In the rest of the paper, we will first introduce in section 2 related work on this topic. We will present next in section 3 the approach we propose. Section 4 will give first experiments and results we have obtained. Conclusion and perspectives arising from this work will be presented in section 5.

## 2 Related work

Performance characterization, in the specific context of localization, is a well known problem in some research topics such as computer vision [6], handwriting segmentation [7], layout analysis [8], text/graphics separation [9], etc. Concerning symbol localization, at the best of our knowledge only the work of [4] has been proposed to

single	an object in the results matches with a single object in the groundtruth
misses	an object in the groundtruth doesn't match with any object in the results
false alarm	an object in the results doesn't match with any object in the groundtruth
multiple	an object in the results matches with multiple objects in the groundtruth (merge case) or an object in the groundtruth matches with multiple objects in the results (split case)

Table 1: Matching cases between groundtruth and localization results

date. The performance characterization algorithms aim to detect possible matching cases between groundtruth and localization results, as detailed in Table 1. Exploiting these results, different criteria could be computed such as segmentation rates or retrieval measures (precision, recall, F-measure).

The key point when developing such a characterization algorithm, is to decide about representations to be used, both in results and groundtruth. Two kinds of approach exist in the literature [8], exploiting pixel-based and geometry-based representations.

In a pixel-based representation, results and groundtruth correspond to sets of pixels. For that reason, algorithms exploiting such a representation are very accurate. They are usually employed to evaluate segmentation tasks in computer vision [6] or handwriting recognition [7]. However, groundtruth creation is more cumbersome and requires a lot more storage. Comparison of groundtruth with the results is also time-consuming.

In a geometry-based representation, algorithms employ geometric shapes to describe regions in both in results and the groundtruth. The type of geometric shapes depend on the application: bounding boxes for text/graphics separation [9], isothetic polygons for layout analysis [8], convex hulls for symbol spotting [4], etc. Comparison of the groundtruth with the results is time-efficient, and the corresponding groundtruth is straightforward to produce. Such a representation is commonly used in the document analysis field, as it is more focused on semantic [9, 8, 4]. Because the main goal of the systems is recognition, evaluation could be limited to detection aspects only (i.e. to decide about a wrong or a correct localization without evaluation of the segmentation accuracy).

In both cases, characterization algorithms exploit information about regions in localization results, and compare them to groundtruth. This results in boolean decisions about positive/negative detections, which raises several open problems:

**Homogeneity of results:** Regions provided as localization results could present a huge variability (set of pixels, bounding boxes, convex hulls, ellipsis, etc.). This variability disturbs the comparison of systems. A characterization algorithm should take these differences into account, and put the results of all the systems at a same level.

**Precision of localization:** Large differences could appear between the size of regions in the results and the groundtruth. These differences correspond to over or under segmentation cases. This results in aberrant detection results. Precision about localization must be considered in the evaluation to compensate it [4].

**Time complexity:** Complex regions, such as symbols, must be represented by pixel sets or polygons to obtain a correct localization precision. However, their comparison is time-consuming [7], and involves to use specific approaches to limit the complexity of the algorithms [8].

In this paper we propose an alternative approach to region-based characterization, to solve these problems. We present this approach in the next section 3.

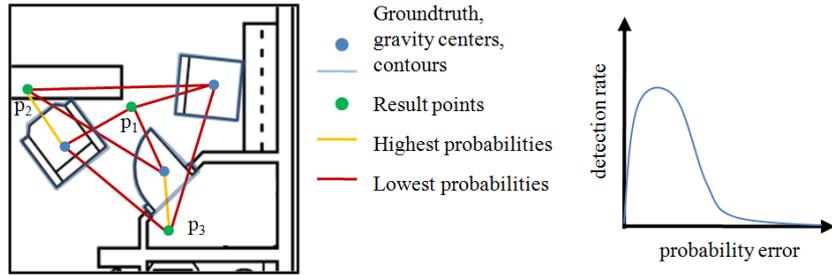


Figure 1: Our approach

### 3 Our approach

#### 3.1 Introduction

Our key objective of this work is to provide a more “generic” and “fuzzy” way to characterize performance of symbol localization systems. To do it, our proposal is to exploit only single points as the results of localization. Indeed, points are a more homogenous and reliable information to evaluate localization. A result could be easily provided as point using the gravity center of the detected region. In addition, because information about the regions in results are not taken into account, this avoids the aberrant detection cases.

However, such an approach makes impossible to detect fully the matching cases between localization results and groundtruth e.g. how to interpret a point slightly outside of a groundtruth region, a point equally distant of two regions, etc. To solve this problem, we exploit in our approach information about context in groundtruth (Fig. 1). For each result point, probability scores are computed with each neighboring regions in the groundtruth. These probability scores will depend on the spatial distribution of the regions in the groundtruth. They will change locally for each result point. Fig. 1 gives an example.  $p_1$ ,  $p_2$  and  $p_3$  are located at similar distances of symbols in groundtruth. However,  $p_2$  and  $p_3$  present highest probability scores, but not  $p_1$ . Local distribution of symbols in groundtruth around  $p_1$  makes the decision about the detection ambiguous. Final characterization results are given with a detection rates/probability error plot, describing the sets of possible interpretations of the localization results.

In this section we describe our characterization algorithm. We exploit three main steps to perform the characterization: (1) in a first step we use a method to compare the localization of a result point to a given symbol in groundtruth (2) exploiting this comparison method, we compute next for each result point its probability scores with all the symbols in groundtruth (3) at last, we employ a matching algorithm to determine the correct detection cases, and draw the detection rates/probability error plots. We will detail each of these steps in next subsections 2.2, 2.3 and 2.4.

#### 3.2 Localization comparison

In our approach, groundtruth is provided as regions (contours with corresponding gravity centers) and localization results as points (Fig. 1). This makes impossible to compare them directly, due to the scaling variations. Indeed, symbols appear at different scales in drawings. To address this problem, we compare the result points with groundtruth regions by computing scale factors (Fig. 2). In geometrical terms, a factor  $s$  specifies how to scale a region in the groundtruth in order to fit its contour with a given result point. Thus, result points inside and outside a symbol will have respectively scale factors of  $s \leq 1$  and  $s > 1$ .

The factor  $s$  is computed from the line  $L$  of direction  $\theta$ , joining the gravity center  $g$  defined in groundtruth to a result point  $r$ . On this line  $L$ ,  $s$  corresponds to the ratio of lengths  $l_{gr}$  and  $l_{gi}$ .  $l_{gr}$  is the Euclidean distance between the gravity center  $g$  and the result point  $r$ .  $l_{gi}$  is computed in the same way, but with the intersection point  $i$  of  $L$  with contours  $c$  of symbol. This intersection point  $i$  is detected using standard line intersection methods [10]. If several intersections exist (i.e. cases of concave contours or holes in symbol), the farthest one

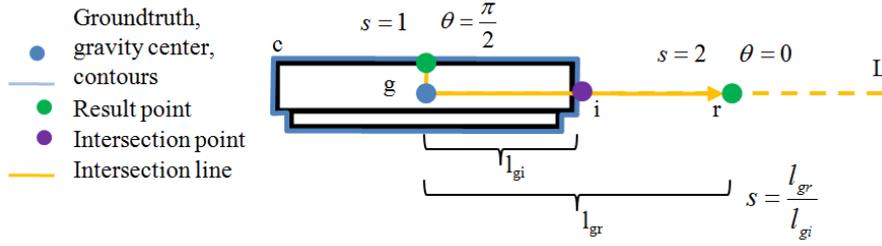


Figure 2: Scale factor

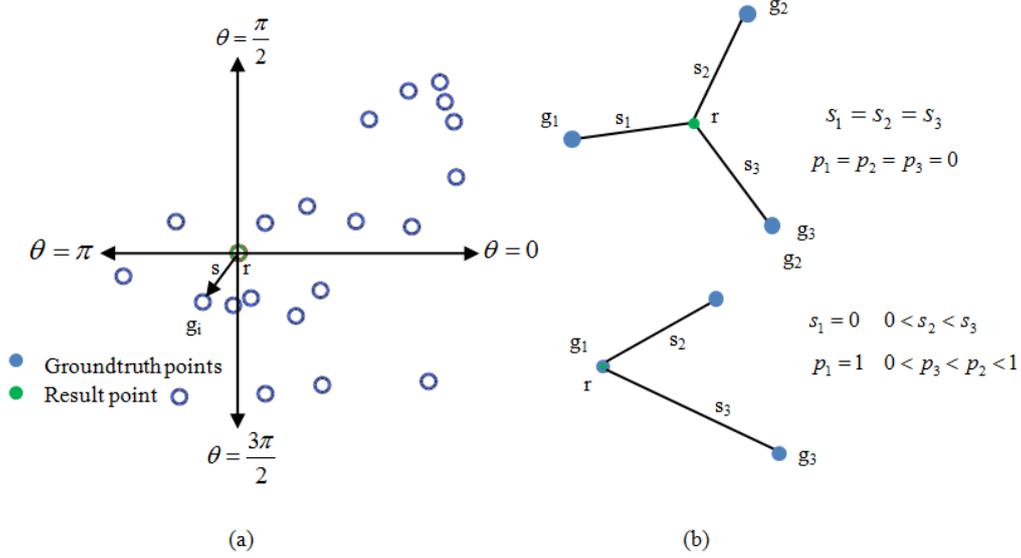


Figure 3: (a) plot  $(\theta, s)$  (b) probability scores

from  $g$  is selected. For further simple implementation of our algorithm, we give also in Appendix how to get  $l_{gi}$  using bounding boxes of symbols.

### 3.3 Probability scores

In a second step, we compute probability scores between result points and groundtruth (Fig. 3). For each result point  $r$ , relations with groundtruth are expressed by a set of  $n$  points  $S = \bigcup_{i=1}^n g_i(\theta, s)$  (Fig. 3 (a)), where  $\theta$  and  $s$  represent respectively the direction and scale factor between the result point and a symbol  $i$ . We define next the probability scores between the result point  $r$  and symbols  $\bigcup_{i=1}^n g_i$  as detailed in Fig. 3 (b). When all symbols  $\bigcup_{i=1}^n g_i$  are equally distant (i.e  $s_1 = s_2 = \dots s_n = s_n$ ), thus  $\bigcup_{i=1}^n p_i = 0$ . In the case of a  $s_i = 0$ , thus the corresponding  $p_i = 1$ . Otherwise, any other cases  $0 < s_i < s_j$  will correspond to probability scores  $0 < p_j < p_i < 1$ .

The equations below give the mathematical details we employ to compute the probability scores. For each  $g_i$ , we compute the probability score  $p_i$  to the result point  $r$  as detailed in (1). To do it, we exploit the other symbols  $\bigcup_{j=1, j \neq i}^n g_j$  in the groundtruth. We mean the values  $f\left(\frac{s_i}{s_j}\right)$  corresponding to local probabilities  $g_i$  to  $r$ , regarding  $g_j$ . The function  $f\left(\frac{s_i}{s_j}\right)$  respects the conditions given in (2). It exists different mathematical ways to define such a function (linear, sinusoidal, etc.). We have employed here a gaussian function (3), as it is a common way to represent random distributions. This function is set using a variance  $\sigma^2 = 1$ , and normalization parameters  $k_y$  and  $k_x$ . These parameters are defined in order to respect the key values  $f(x = 0) = 1$  and

$f(x \geq 1) = 0$ . To respect the second condition, we bound to 0 the gaussian values  $f(x \geq 1)$ , with a  $k_x$  determined to respect a threshold error  $\lambda^1$ .

$$p_i = \sum_{j=1, j \neq i}^n \frac{f\left(\frac{s_i}{s_j}\right)}{n-1} \quad (1)$$

$$s_i = 0 \rightarrow f\left(\frac{s_i}{s_j} = 0\right) = 1 \quad (2)$$

$$s_i = s_j \rightarrow f\left(\frac{s_i}{s_j} = 1\right) = 0$$

$$s_i \rightarrow +\infty \rightarrow f\left(\frac{s_i}{s_j} \rightarrow +\infty\right) = 0$$

$$s_j = 0 \rightarrow f\left(\frac{s_i}{s_j} \rightarrow +\infty\right) = 0$$

$$s_j \rightarrow +\infty \rightarrow f\left(\frac{s_i}{s_j} \rightarrow 0\right) \rightarrow 1$$

$$f\left(x = \frac{s_i}{s_j}\right) = \frac{k_y}{\sqrt{2\pi}} \times \exp^{-\frac{(k_x x)^2}{2}} \quad (3)$$

$$k_y = \sqrt{2\pi}$$

$$\frac{\sqrt{2\pi}}{2} - \int_0^1 f(x) dx < \lambda \rightarrow k_x$$

### 3.4 Matching algorithm

Once probability scores are computed, we look for relevant matchings between groundtruth and results of a system. We propose here a ‘‘fuzzy’’ way to find these matchings. We compute different sets of matching results, according to the ranges of the probability scores. Final results are displayed into plots (Fig. 4 (a)), where the x axis corresponds to score error  $\varepsilon$  (i.e. inverse of probability score  $\varepsilon = 1 - p$ ), and the y axis to performance rates.

We compute different performance rates,  $T_s$ ,  $T_f$  and  $T_m$  corresponding respectively to single detections, false alarms and multiple detections (see Table 1). The rate of ‘‘misses cases’’ corresponds to  $1 - T_s$ . To do it, we build-up a correspondence list between results and groundtruth as detailed in Fig. 4 (b). This correspondence list is bi-partite, composed of nodes corresponding to the groundtruth  $\bigcup_{i=1}^n g_i$  and the results  $\bigcup_{j=1}^q r_j$ . Our probability scores are given as undirected arcs  $\bigcup_{k=1}^{np} a_k = (g_i, r_j, p_{ij})$  of  $np$  size. We use these arcs to make the correspondences in the list in an incremental way, by shifting the  $\varepsilon$  value from 0 to 1. An arc  $a_k$  is added to the list when its  $p_{ij} < \varepsilon$ . For each  $\varepsilon$  value, the  $T_s$  and  $T_f$  rates are computed by browsing the list, and checking the degrees of nodes  $\bigcup_{i=1}^n d_{g_i}$  and  $\bigcup_{j=1}^q d_{r_j}$ , as detailed in (4), (5) and (6).

$$\forall g_i \leftrightarrow r_j, d_{g_i} = d_{r_j} = 1 \rightarrow s = s + 1 \quad (4)$$

$$T_s = \frac{s}{n}$$

$$\forall r_i, d_{r_i} = 0 \rightarrow f = f + 1 \quad (5)$$

$$T_f = \frac{f}{q}$$

<sup>1</sup>In our algorithm,  $\lambda$  is fixed at  $\sqrt{2\pi}10^{-5}$  corresponding to  $k_x = 3.9$ .

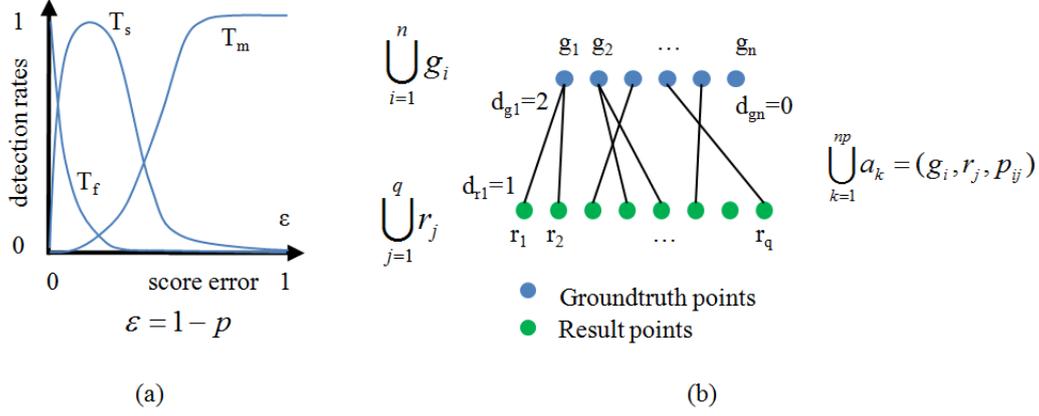


Figure 4: (a) result plot (b) correspondence list

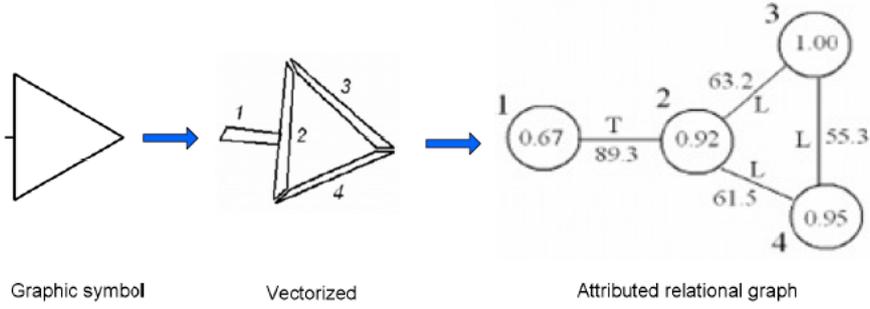


Figure 5: Representation phase of [1]

$$\forall r_j \leftrightarrow g_i, d_{r_j} > 1 \vee d_{g_i} > 1 \rightarrow m = m + 1 \quad (6)$$

$$T_m = \frac{m}{q}$$

## 4 Experiments and Results

In this section, we present a set of preliminar experiments and results obtained using our algorithm. We have applied it to evaluate the symbol localization system of [1]. This system relies on a structural approach, using a two-step process.

First, it extracts topological and geometric features from an image, and structures them through an ARG<sup>2</sup> (Fig. 5). The image is preliminary vectorized into a set of quadrilateral primitives. These primitives become nodes in the ARG (labels 1, 2, 3, 4), and connection between them arcs. Nodes have, as attributes, relative lengths (normalized between 0 and 1) whereas arcs have connection-type (*L junction, T junction*) and relative angle (normalized between 0° and 90°).

In the second step, the system looks for potential regions of interest corresponding to symbols in the image. It detects parts of the graph that may correspond to symbols i.e. symbol seeds. Scores, corresponding to probabilities of being part of a symbol, are computed for all edges and nodes of the graph. They are based on features such as lengths of segments, perpendicular and parallel angular relations, degrees of nodes, etc. The symbol seeds are detected next during a score propagation process. This process seeks and analyses the different shortest paths and loops between nodes in the graph. The scores of all the nodes belonging to a detected path

<sup>2</sup>Attributed Relational Graph

	Floorplan level		Symbol level	
<b>Setting</b>	backgrounds	5	models	16
<b>Dataset</b>	images	100	symbols	2521

Table 2: Dataset used for experiments

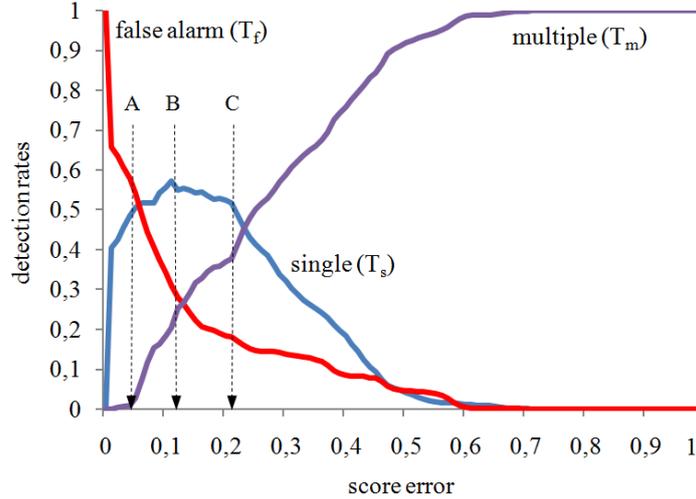


Figure 6: Characterization results

are homogenized (propagation of the maximum score to all the nodes in the path) until convergence, to obtain the seeds.

To test this system, we have used test documents coming from the SESYD database<sup>3</sup>. This database is composed of synthetical documents, with the corresponding groundtruth, produced using the system described in [3]. This system allows the generation of synthetic graphics documents, containing non-isolated symbols in a real context. It is based on the definition of a set of constraints, that permit to place the symbols on a predefined background, according to the properties of a particular domain (architecture, electronics, engineering, etc.). The SESYD database is composed of different collections, including architectural floorplans, electrical diagrams, maps, etc. In this work, we have limited our experiments to architectural floorplans. The Table 2 gives details about the dataset we have used.

The next Fig. 6 presents the characterization results we have obtained for these experiments. The system presents in *A* a good confidence for the detection results  $T_s \leq 0.504$ , with an score error  $\varepsilon \leq 0.05$  and nearly none multiple detections  $T_m \leq 0.027$ . However, these detection results are joined to a  $T_f = 0.542$ , highlighting a bad precision of the system. The best detection results are obtained in *B* with a  $T_s = 0.571$ , corresponding to an score error of  $\varepsilon = 0.11$ . However, at this state confusions appear in localization results with a  $T_m \geq 0.204$ . This rate results in the merging of false alarms with less confident results, as the false alarm rate goes down to  $T_f = 0.309$ . Certainly the systems will increase in a significative way its detection results  $T_s$ , by improving its precision. Up to this point,  $T_s$  dies down slowly, and for score error  $\varepsilon \geq 0.21$  in *C* the  $T_s$  and  $T_m$  curves start to be diametrically opposite.

<sup>3</sup><http://mathieu.delalandre.free.fr/projects/sesyd/>

## 5 Conclusions and Perspectives

In this paper we have presented an algorithm for performance characterization of symbol localization systems. This work aims to propose a more “generic” and “fuzzy” way to characterize performance. Our proposal is to exploit only single points as the results of localization, as it is a more homogenous and reliable information to evaluate localization. To compare them with the groundtruth, we exploit information about context. For each result point, probability scores are computed with each neighboring regions in the groundtruth. These probability scores will depend on the spatial distribution of the regions in the groundtruth. They will change locally for each result point. Final characterization results are given with a detection rates/probability error plot, describing the sets of possible interpretations of the localization results. We present preliminar experiments and results obtained using our algorithm, to evaluate the symbol localization system of [1]. These experiments have been done on a dataset of synthetic floorplans (with the corresponding groundtruth), composed of 100 of document images and around 2500 symbols. We conclude about the performance of this system, in terms of localization accuracy and precision level (false alarms and multiple detections).

Concerning the perspectives of this work, our short term ones are to extend our experiments to noisy images [11] (to learn about the robustness of the system) and scalable datasets (with more symbol models, or from other application domains “electrical diagrams, geographic maps, etc.”). We plan also to perform experiments with real datasets [4], and to compare the obtained results with the synthetic ones. Concerning the improvements of our algorithm, we plan to investigate ways to compare automatically the curves obtained by different localization systems, as the error goes from 0 to 1. At last, our long term perspective concerns the comparison of different symbol localization systems. We plan to take benefit of the work done around the organization of the international symbol recognition contest 2009<sup>4</sup>, to look for potential participants interested in testing their systems with this characterization approach.

## 6 Acknowledgements

This work has been supported by the Spanish projects TIN2008-04998 and Consolider Ingenio 2010 (CSD2007-00018). The authors wish to thank Hervé Locteau (LORIA institute, Nancy, France) for his comments and corrections about the paper.

## References

- [1] R. Qureshi, J. Ramel, D. Barret, H. Cardot, Symbol spotting in graphical documents using graph representations, in: Workshop on Graphics Recognition (GREC), Vol. 5046 of Lecture Notes in Computer Science (LNCS), 2008, pp. 91–103.
- [2] M. Delalandre, T. Pridmore, E. Valveny, E. Trupin, H. Locteau, Building synthetic graphical documents for performance evaluation, in: Workshop on Graphics Recognition (GREC), Vol. 5046 of Lecture Note in Computer Science (LNCS), 2008, pp. 288–298.
- [3] M. Delalandre, E. Valveny, J. Lladós, Performance evaluation of symbol recognition and spotting systems: An overview, in: Workshop on Document Analysis Systems (DAS), 2008, pp. 497–505.
- [4] M. Rusiñol, J. Lladós, A performance evaluation protocol for symbol spotting systems in terms of recognition and location indices, International Journal on Document Analysis and Recognition (IJ DAR).
- [5] H. Locteau, S. Adam, E. Trupin, J. Labiche, P. Heroux, Symbol spotting using full visibility graph representation, in: Workshop on Graphics Recognition (GREC), 2007, pp. 49–50.

---

<sup>4</sup><http://epeires.loria.fr/public/general09>

- [6] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, *Pattern Analysis and Machine Intelligence (PAMI)* 29 (6) (2007) 929–944.
- [7] T. Breuel, Representations and metrics for off-line handwriting segmentation, in: *International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, 2002, pp. 428–433.
- [8] D. Bridson, A. Antonacopoulos, A geometric approach for accurate and efficient performance evaluation, in: *International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.
- [9] L. Wenyin, D. Dori, A proposed scheme for performance evaluation of graphics/text separation algorithms, in: *Workshop on Graphics Recognition (GREC)*, Vol. 1389 of *Lecture Notes in Computer Science (LNCS)*, 1997, pp. 335–346.
- [10] I. Balaban, An optimal algorithm for finding segments intersections, in: *Symposium on Computational Geometry (SGC)*, 1995, pp. 211–219.
- [11] T. Kanungo, R. M. Haralick, I. Phillips, Non-linear local and global document degradation models, *International Journal of Imaging Systems and Technology (IJIST)* 5 (3) (1994) 220–230.

## Appendix

