A Performance Characterization Algorithm for Symbol Localization

Mathieu Delalandre[1,2], Jean-Yves Ramel[1], Ernest Valveny[2], and Muhammad Muzzamil Luqman[1,2]

[1] Laboratoire d'Informatique (LI), 37200 Tours, France
mathieu.delalandre@univ-tours.fr
jean-yves.ramel@univ-tours.fr
muhammadmuzzamil.luqman@etu.univ-tours.fr
[2] Computer Vision Center (CVC), 08193 Bellaterra (Barcelona), Spain
ernest@cvc.uab.es

Abstract. In this paper we present an algorithm for performance characterization of symbol localization systems. This algorithm is aimed to be a more "reliable" and "open" solution to characterize the performance. To achieve that, it exploits only single points as the result of localization and offers the possibility to reconsider the localization results provided by a system. We use the information about context in groundtruth, and overall localization results, to detect the ambiguous localization results. A probability score is computed for each matching between a localization point and a groundtruth region, depending on the spatial distribution of the other regions in the groundtruth. Final characterization is given with detection rate/probability score plots, describing the sets of possible interpretations of the localization results, according to a given confidence rate. We present experimentation details along with the results for the symbol localization system of [1], exploiting a synthetic dataset of architectural floorplans and electrical diagrams (composed of 200 images and 3861 symbols).

Keywords: symbol localization, groundtruth, performance characterization

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 [4,5], and different systems working at localization level in [6,1,4]. 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 symbols provided by a given system, and the corresponding ones 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 experiments and results we have obtained with our algorithm. 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 [7], handwriting segmentation [8], layout analysis [9], text/graphics separation [10], etc. Concerning symbol localization, at the best of our knowledge only the work of [4] has been proposed to date. Performance characterization algorithms (in the specific context of localization) aim to detect possible matching cases between groundtruth and localization results, as detailed in Table 1. They determine about true or false localization results, without considering the class of objects. It is therefore a two-class recognition problem, to separate objects from background. Once objects are correctly located/segmented, we could proceed the evaluation of recognition. This is known as whitebox evaluation in the literature [11], the goal is to characterize the performance of individual submodules of a complete system and to see how they interact each other.

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

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 [9], 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 [7] or handwriting recognition [8]. 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 the regions (in results and groundtruth). The type of geometric shapes depend on the application: bounding boxes - text/graphics separation [10], isothetic polygons - layout analysis [9], convex hulls - symbol spotting [4], etc. Comparison of the groundtruth

with the results is time-efficient, and the corresponding groundtruth straightforward to produce. Such a representation is commonly used in the document analysis field, as it is more focused on semantic [10,9,4]. 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.

Reliability of results: 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 positive matching cases between the groundtruth and the detection results, when large regions in the results intersects smallest ones in groundtruth. To be able to detect these ambiguous cases a characterization algorithm should be able to reconsider the localization results provided by a system.

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 both for geometry-based and pixel-based representations [9,8]. This involves to use specific approaches to limit the complexity of the algorithms [9].

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.

3 Our approach

3.1 Introduction

Our objective in this work is to provide a "reliable" and "open" solution to characterize the performance of symbol localization systems. To achieve that, we propose an algorithm exploiting only single points as the result of localization and offering the possibility to reconsider the localization results provided by a system. It uses the information about context in groundtruth, and overall localization results, to detect the ambiguous localization cases. A probability score is computed for each matching between a localization point and a groundtruth region, depending on the spatial distribution of the other regions in the groundtruth. Final characterization is given with detection rate/probability score plots, describing the sets of possible interpretations of the localization results, according to a given confidence rate.

Fig.1 illustrates our approach. For each result point r_i, probability scores are computed with each symbol g_i in the groundtruth. These probability scores will depend on the spatial distribution of the symbols g_i in the groundtruth, they will change locally

for each result point r_i. In Fig. 1, r_1, r_2 and r_3 are located at similar distances of symbols g_1, g_2 and g_3 in groundtruth. However, r_2 and r_3 present highest probability scores to be matched with g_2 g_3 respectively, but not r_1. Local distribution of symbols g_i in groundtruth around r_1 makes ambiguous the characterization of this localization result. For that reason, a positive matching between r_1 and any symbols g_i will be considered with a low probability. Final characterization is given with a detection rate/probability score plot, describing the sets of possible interpretations of the localization results according to their probability scores.

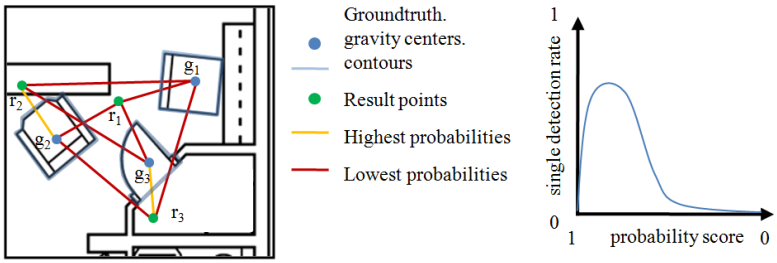


Fig. 1. Our approach

In the rest of 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 identify the correct detection cases, and draw the detection rate/probability score plots. We will detail each of these steps in next subsections 3.2, 3.3 and 3.4 respectively. Table 2. provides the list of mathematical symbols that we have used for detailing the different steps of our proposed algorithm.

i	index value	j	index value
r_i	result object	g_i	groundtruth object
θ	direction between result and groundtruth objects	s_i	scaling factor
q	number of result objects	n	number of groundtruth objects
p_i	probability score	$f(s_i/s_j)$	probability score function
ε	score error ($\varepsilon = 1 - p_i$)		
s	single detection	T_s	single detection rate (=s/n)
f	false detection	T_f	false alarm rate (=f/q)
m	multiple detection	T_m	multiple detection rate (=m/q)

Table 2. Table of mathematical symbols

3.2 Localization comparison

In our approach, groundtruth is provided as regions (contours with corresponding gravity centers) and localization results as points (Fig. 1). We ask the systems to provide points that should be in regions near the location of symbols in groundtruth. These points could be gravity centers, key-points of interest or junctions. The use of points as results makes impossible to compare them directly with groundtruth, 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 so that its contour fits with a given result point. Thus, result points inside and outside a symbol will have respectively scale factors of $s \leq 1$ and $s > 1$.

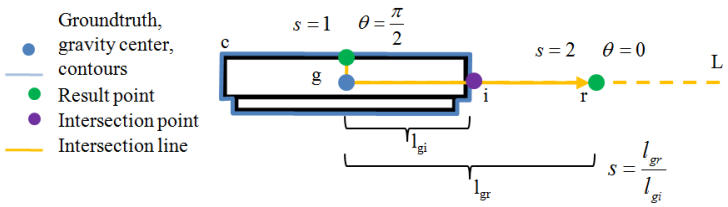


Fig. 2. Scale factor

The factor s is computed from the line L of direction θ, 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 [12]. If several intersections exist (i.e. cases of concave contours or holes in symbol), the farthest one from g is selected.

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 θ 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 $s_1 = s_2 = ...s_i.. = s_n$), 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 (**1**) (**2**) 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

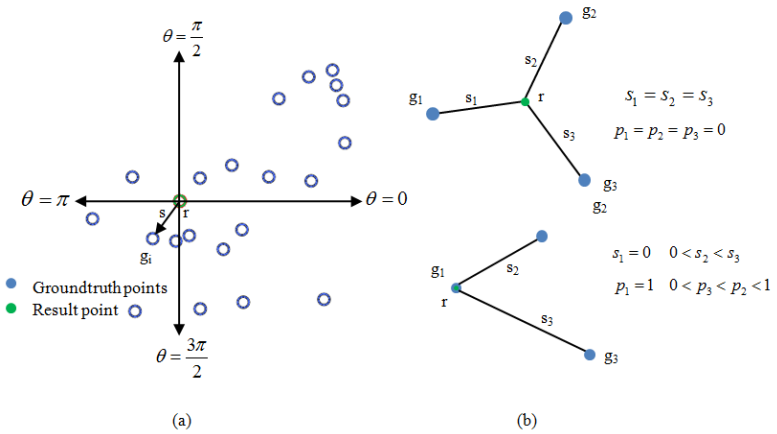


Fig. 3. (a) plot (θ, s) (b) probability scores

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 $f\left(\frac{s_i}{s_j}\right)$ must respect the conditions given in Table 3. In fact, there exists different mathematical ways to define such a function (based on inverse, linear, cosinus functions, etc.). We have employed here a gaussian function (**2**), 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 obtain the key values $f(x = 0) = 1$ and $f(x = 1) \rightarrow 0$. Fig. 4 gives plots of gaussian and probability score functions to illustrate how kx and ky impact the values. The value of k_y is defined as $\sqrt{2\pi}$, to bound $f(x = 0)$ at 1 (e^0). Concerning k_x, we determine it using the approximation into a taylor serie of the cumulative distribution function $\Phi_0(x) = \int_0^x \phi(x)$. We define k_x when $\Phi_0(x) = 1 - \lambda$, with λ a precision defined manually. In our experiments, we have fixed $\lambda = 10^{-6}$ corresponding to $kx = 3.9$.

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

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

$f\left(x = s_i / s_j\right)$	0		1	$+\infty$	
	$s_i = 0$	$s_i \rightarrow +\infty$	$s_i = s_i$	$s_i \rightarrow +\infty$	$s_i \rightarrow 0$

Table 3. Table of function $f(x = \frac{s_i}{s_j})$

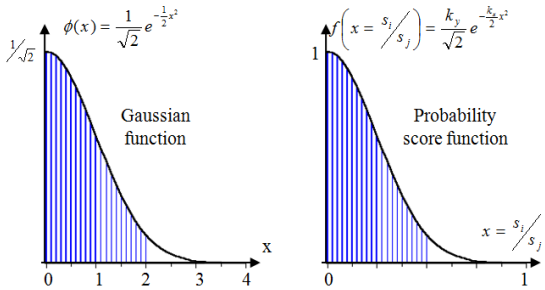


Fig. 4. Gaussian and probability score functions

3.4 Matching algorithm

Once probability scores are computed, we look for relevant matchings between groundtruth and results of a system. In order to make open interpretation of localization of symbols, we provide characterization results using distribution plots. We compute different sets of matching results, according to the ranges of the probability scores. Final results are displayed into distribution plots (Fig. 5 (a)), where the x axis corresponds to score error ε (i.e. inverse of probability score $\varepsilon = 1 - p$), and the y axis to performance rates.

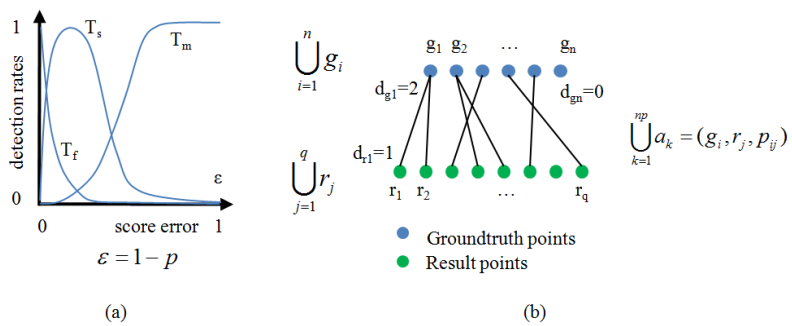


Fig. 5. (a) result plot (b) correspondence list

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. 5 (b). This correspondence list is bipartite, 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 $\bigcup_{k=1}^{nq} a_k = (g_i, r_j, p_{ij})$ of nq size. We use these arcs to make the correspondences in the list in an incremental way,

by shifting the ε value from 0 to 1. An arc a_k is added to the list when its $p_{ij} \geq 1 - \varepsilon$. For each ε value, the T_s and T_f rates are computed by browsing the list, and checking the degrees of nodes $\bigcup_{i=1}^{n} d_{gi}$ and $\bigcup_{j=1}^{q} d_{rj}$, as detailed in (3), (4) and (5).

$$\forall g_i \leftrightarrow r_j, d_{gi} = d_{rj} = 1 \rightarrow s = s + 1 \quad (3)$$
$$T_s = \frac{s}{n}$$
$$\forall r_j, d_{rj} = 0 \rightarrow f = f + 1 \quad (4)$$
$$T_f = \frac{f}{q}$$
$$\forall r_j \leftrightarrow g_i, d_{rj} > 1 \vee d_{gi} > 1 \rightarrow m = m + 1 \quad (5)$$
$$T_m = \frac{m}{q}$$

4 Experiments and Results

In this section, we present 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 a given image, and represents it using an ARG[3] (Fig. 6). The image is preliminary vectorized into a set of quadrilateral primitives. These primitives become nodes in the ARG (labels 1, 2, 3, 4), and connections between become arcs. Nodes have, as attributes, relative lengths (normalized between 0 and 1) whereas arcs have connection-type (*L junction*,*T junction*, *X junction*, etc.) and relative angle (normalized between 0° and 90°).

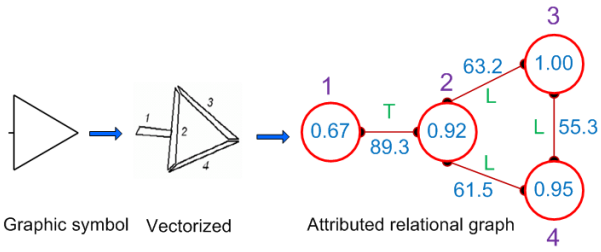


Fig. 6. Representation phase of [1]

In the second step, the system looks for potential regions of interest corresponding to symbols. It detects parts of the ARG 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 ARG. They are based on features such as lengths of segments,

[3] Attributed Relational Graph

perpendicular and parallel angular relations, degrees of nodes, etc. The symbol seeds are detected next during a score propagation process. This process seeks and analyzes the different shortest paths and loops between nodes in the ARG. The scores of all the nodes belonging to a detected path 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 employed datasets coming from the SESYD database[4]. This database is composed of synthetic document images, with the corresponding ground-truth, produced using the system described in [5]. This system allows the generation of synthetic graphic 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 pre-defined background, according to the properties of a particular domain (architecture, electronics, etc.). The SESYD database is composed of different collections, including architectural floorplans, electrical diagrams, geographic maps, etc. In this work, we have limited our experiments to subsets of this database, including documents from the electrical and architectural domains. Table 4. gives details of the dataset we use, and Fig. 7 gives some examples of test documents.

		Drawing level		Symbol level	
floorplans	**Setting**	backgrounds	5	models	16
	Dataset	images	100	symbols	2521
diagrams	**Setting**	backgrounds	5	models	17
	Dataset	images	100	symbols	1340

Table 4. Dataset used for experiments

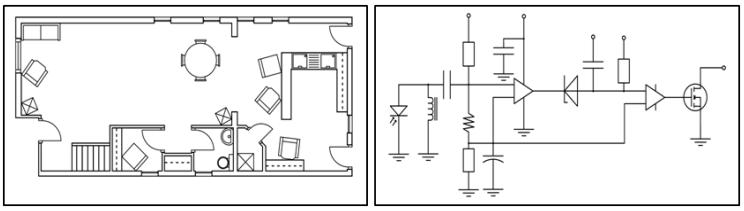
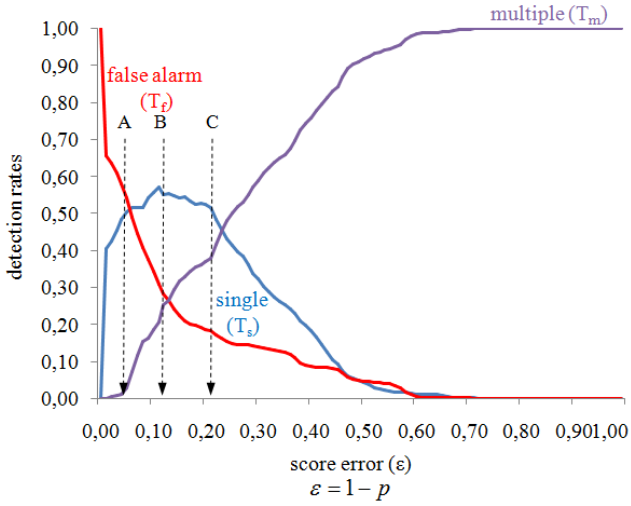


Fig. 7. Examples of test documents

Fig. 8 presents the characterization results we have obtained on floorplans, with variation of $\{T_s, T_f, T_m\}$ rates. The system presents in A a good confidence for the detection results $T_s \leq 0.50$, with an score error $\varepsilon \leq 0.05$ and nearly none multiple

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

detections $T_m \leq 0.03$. The best detection results is obtained in B with a $T_s = 0.57$, corresponding to an score error of $\varepsilon = 0.11$. However, these detection results are joined to a $T_f = 0.31$, highlighting a bad precision of the system. In addition, at this state confusions appear in localization results with a $T_m \geq 0.20$. This rate results in the merging of false alarms with less confident results, as the false alarm rate goes down to $T_f = 0.31$. 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.



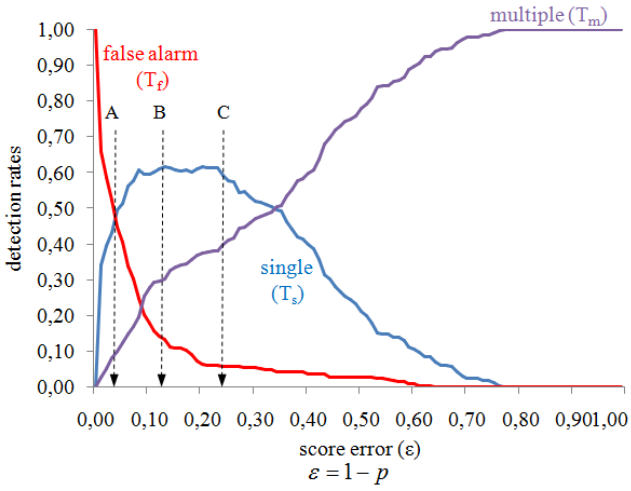
floorplans

	A	B	C
ε	0.05	0.11	0.21
T_s	0.50	0.57	0.52
T_f	0.54	0.31	0.18
T_m	0.03	0.20	0.38

Fig. 8. Characterization results on floorplans

Fig. 9 gives results concerning electrical diagrams. The system presents a good confidence in A, for $\varepsilon \leq 0.04$ and $T_s = T_f = 0.45$. However, at this point the system already does multiple detections with a $T_m = 0.10$. The best localization score is obtained at B with $\varepsilon = 0.13$ and $T_s = 0.62$. Therefore, the best localization score is better for electrical diagrams than floorplans. In addition, the system doesn't make a lot of false detections with a $T_f = 0.13$. However, multiple detections stay higher with $T_m = 0.30$. Up to C, T_s decreases in linear way for score errors $\varepsilon \geq 0.23$.

When comparing results obtained by the system [1] on these two application domains, the ones on electrical diagrams are best. However, each of them illustrates specific failures of the system. The first is the high level of multiple detection appearing on electrical diagrams. Certainly the systems could increase a lot its detection results, by introducing split / merge procedures of detected regions of interest. The second concerns the generated false alarms on floorplans. In order to reduce this problem, a point to be improved in the system is to introduce a checking procedure of detected ROIs, to reduce the hypothesis of localization.



electrical diagrams

	A	B	C
ε	0.04	0.13	0.23
T_s	0.45	0.62	0.61
T_f	0.45	0.13	0.06
T_m	0.10	0.30	0.38

Fig. 9. Characterization results on electrical diagrams

5 Conclusions and Perspectives

In this paper we have presented an algorithm for performance characterization of object localization systems. This algorithm has been applied in the context of symbol localization, but we believe it could be applied to other problems (medical image segmentation, mathematical formula recognition . . .). It aims to propose a more "reliable" and "open" solution to characterize the performance of systems, by offering the possibility to reconsider the localization results. It exploits only single points as the results of localization. Then a probability score is computed for each matching between a localization point and a groundtruth region, depending on the spatial distribution of the other regions in the groundtruth. They will change locally for each result point. Characterization results are given with detection rate/probability score plots, describing the sets of possible interpretations of the localization results, according to a given confidence rate. We present 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 images (with the corresponding groundtruth), composed of 200 document images and around 3861 symbols from electrical and architectural domains. We conclude about the performance of this system, in terms of localization accuracy, precision level (false alarms and multiple detections) on both datasets.

In future, we aim to take-forward our experimentations to evaluate the scalability (large number of symbol models) and robustness (noisy images [13]) of systems. We also plan to perform experiments with real datasets [4], and to compare the obtained results with the synthetic ones. And, our final goal is the comparison of different symbol localization systems. We plan to take benefit of the work done around the EPEIRES

project[5], to look for potential participants interested in testing their systems with this characterization approach.

6 Acknowledgements

This work has been partially supported by the Spanish projects CONSOLIDER-INGENIO 2010 (CSD2007-00018), TIN2008-04998 and TIN2009-14633-C03-03. The authors wish to thank Hervé Locteau (LORIA institute, Nancy, France) for his comments and corrections about the paper.

References

1. Qureshi, R., Ramel, J., Barret, D., Cardot, H.: Symbol spotting in graphical documents using graph representations. In: Workshop on Graphics Recognition (GREC). Volume 5046 of Lecture Notes in Computer Science (LNCS). (2008) 91–103
2. Valveny, E., Tabbone, S., Ramos, O., Philippot, E.: Performance characterization of shape descriptors for symbol representation. In: Workshop on Graphics Recognition (GREC). Volume 5046 of Lecture Notes in Computer Science (LNCS). (2008) 278–287
3. Delalandre, M., Valveny, E., Lladós, J.: Performance evaluation of symbol recognition and spotting systems: An overview. In: Workshop on Document Analysis Systems (DAS). (2008) 497–505
4. 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 (IJDAR) **12**(2) (2009) 83–96
5. Delalandre, M., Pridmore, T., Valveny, E., Trupin, E., Locteau, H.: Building synthetic graphical documents for performance evaluation. In: Workshop on Graphics Recognition (GREC). Volume 5046 of Lecture Note in Computer Science (LNCS). (2008) 288–298
6. Locteau, H., Adam, S., Trupin, E., Labiche, J., Heroux, P.: Symbol spotting using full visibility graph representation. In: Workshop on Graphics Recognition (GREC). (2007) 49–50
7. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. Pattern Analysis and Machine Intelligence (PAMI) **29**(6) (2007) 929–944
8. Breuel, T.: Representations and metrics for off-line handwriting segmentation. In: International Workshop on Frontiers in Handwriting Recognition (IWFHR). (2002) 428–433
9. Bridson, D., Antonacopoulos, A.: A geometric approach for accurate and efficient performance evaluation. In: International Conference on Pattern Recognition (ICPR). (2008) 1–4
10. Wenyin, L., Dori, D.: 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). (1997) 335–346
11. Kanungo, T., Resnik, P.: The bible, truth, and multilingual ocr evaluation. In: Document Recognition and Retrieval (DRR). Volume 3651 of SPIE Proceedings. (1999) 86–96
12. Balaban, I.: An optimal algorithm for finding segments intersections. In: Symposium on Computational Geometry (SGC). (1995) 211–219
13. Kanungo, T., Haralick, R.M., Phillips, I.: Non-linear local and global document degradation models. International Journal of Imaging Systems and Technology (IJIST) **5**(3) (1994) 220–230

[5] http://epeires.loria.fr/