

Spatio-structural Symbol Description with Statistical Feature Add-on

Santosh K.C., Bart Lamiroy, Laurent Wendling

► **To cite this version:**

Santosh K.C., Bart Lamiroy, Laurent Wendling. Spatio-structural Symbol Description with Statistical Feature Add-on. Workshop on Graphics RECOgnition (GREC), Sep 2011, Seoul, South Korea. 2011. <inria-00617296>

HAL Id: inria-00617296

<https://hal.inria.fr/inria-00617296>

Submitted on 26 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatio-structural Symbol Description with Statistical Feature Add-on

Santosh K.C.*, Bart Lamiroy[†] and Laurent Wendling[‡]

*INRIA Nancy-Grand Est, [†]Nancy Université INPL – LORIA, [‡]LIPADE, Université Paris Descartes
 Email: Santosh.KC@inria.fr; Bart.Lamiroy@loria.fr; Laurent.Wendling@parisdescartes.fr

Abstract—In this paper, we present a method for symbol description based on spatio-structural as well as statistical features of visual elementary parts called ‘vocabulary’. The extracted vocabulary is first organised into different groups based on their types (e.g., circle, corner). This vocabulary is used as a basis for an Attributed Relational Graph (ARG) where spatial relational descriptors formalise the links between the types, labelled with global shape descriptors. The description is used to globally recognise structure by comparing the signatures. The method is experimentally validated in the context of electrical symbol recognition from wiring diagrams.

Keywords—Graphical Symbol; Vocabulary; Spatial Relations; ARG; Shape Descriptors.

I. INTRODUCTION

Graphics recognition has an extremely rich state-of-the-art literature in symbol recognition and localisation. However, most methods are particularly suited for isolated line symbols, not for composed symbols connected to a complex environment [4], [10]. Considering the problem of symbol localisation in real documents, composed of individual parts constrained by spatial relations for instance, one needs to be able to extract visual parts and formalise the possible links that exist between them, in addition to the possible use of shape description on visual parts. This integration of spatial relations and shape description of the extracted visual parts is going to be the core of this paper. The method is very much inspired by a real world industrial problem [16], [7]. Fig. 1 shows a few samples of it.

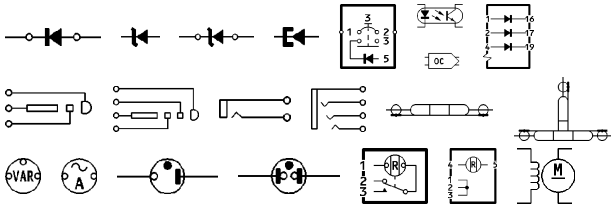


Figure 1. A few symbols in FRESH dataset. Both linear as well as symbols in the composite form are enlisted.

Global signal based descriptors [21] present a number of inconvenients in our context. They difficultly accommodate with connected as well as composite symbols. They are generally not well adapted for capturing small detail changes. In statistical approaches, signatures are simple with low computational cost. However, discriminative power and robustness

strongly depend on the optimal selection of features [1] as well as fusion of classifiers [14] partially based on the idea presented in [15] that off-the-shelf methods are primarily designed for applications where line symbols are isolated. Besides global signal based descriptors, another idea is to decompose the symbols into either vector based primitives like *points*, *lines* and *arcs* or, on the other hand, meaningful parts like *circles*, *triangles* and *rectangles*. They are then represented as Attributed Relational Graphs (ARG) [3], Region Adjacency Graphs (RAG) [9], as well as deformable templates [17]. In addition to the common drawback *i.e.*, error-prone raster-to-vector conversion, variability of the size of graphs leads to computational complexity in matching. Structural approaches however, provide a powerful relational representation, conveying how parts are connected to each other, while also preserving generality and extensibility.

In this paper, we aim to combine the best of both structural and statistical approaches, and try to avoid the shortcomings of each of them. To do so, we decompose symbols by expressing their various parts in a fixed visual vocabulary, using spatial relations, graphs and signal based descriptors to describe the whole shape. Our symbol description is clearly explained in Section II. Thanks to the use of fixed and completely labelled attributes, we can avoid the NP-hardness of the matching problem. Overall, symbols can be compared by fusion of matching scores from every corresponding vertex and edge alignment (*cf.*, Section III). In Section IV, we validate that our method. The paper concludes in Section V.

II. SYMBOL DESCRIPTION

We have a set of well controlled visual elementary parts as a *vocabulary* [16], [7] (see Fig. 2). We denote the type set as, $\sum_{\mathbb{T}} = \{\mathbb{T}_{thick}, \mathbb{T}_{circle}, \mathbb{T}_{corner}, \mathbb{T}_{extremity}\}$.

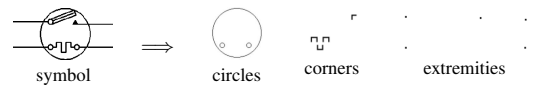


Figure 2. Visual elementary parts from corresponding symbol.

Now, we represent whole symbol by a complete ARG as a 4-tuple $G = (V, E, F_A, F_E)$ where V is the set of vertices, $E \subseteq V \times V$ is the set of graph edges, $F_A : V \rightarrow A_V$ is a function assigning attributes to the vertices and A_V representing a set of vocabulary type set $\sum_{\mathbb{T}}$ as well as their global shape signatures S , and $F_E : E \rightarrow \mathbb{R}_E$ is a function assigning labels to the edges where \mathbb{R} represents spatial

relations of the edge E . Following Fig. 2 and its resulting graph in Fig. 3 whose attribute type set is $\{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3\}$, its ARG representation can be expressed as $G = \{$

$$\begin{aligned} V &= \{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3\}, E = \{(\mathbb{T}_1, \mathbb{T}_2), (\mathbb{T}_1, \mathbb{T}_3), (\mathbb{T}_2, \mathbb{T}_3)\}, \\ F_A &= \{((\mathbb{T}_1, \mathbb{T}_{circle}), \mathcal{S}(\mathbb{T}_1)), \dots, ((\mathbb{T}_3, \mathbb{T}_{extremity}), \mathcal{S}(\mathbb{T}_3))\} \\ F_E &= \{((\mathbb{T}_1, \mathbb{T}_2), \mathfrak{R}(\mathbb{T}_1, \mathbb{T}_2)), \dots, ((\mathbb{T}_2, \mathbb{T}_3), \mathfrak{R}(\mathbb{T}_2, \mathbb{T}_3))\}. \end{aligned}$$

Since this forms a complete graph, it is obvious that there exist $r = \frac{t(t-1)}{2}$ edges for t attribute types.

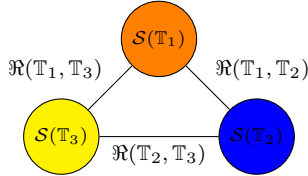


Figure 3. ARG description for a symbol in Fig. 2 – an example.

A. Edge Signature

Choice of spatial relation models rely on behaviour of the studied objects. If the objects are far enough from each other, their relations can be approximated by their centres based on the discretised angle [11]. Otherwise, if they are neither too far nor too close, relations can be approximated by their Minimum Bounding Rectangle (MBR) [12] as long as they are regular. Approaches like histograms of angle [18], tends to be more capable of dealing with overlapping. However, since they consider all pixels, their computational cost increases dramatically. Our work is inspired by the concept of fuzzy relations that takes degree of truth which is more natural than using standard, all-or-none relations [5].

Our model can be explained as follows. For a given reference point \mathbb{R}_p , we cover the surrounding space at regular radial intervals of $\Theta = 2\pi/m$ by using a radial-line. We set up a unique reference point \mathbb{R}_p computed from centroids \mathbb{C}_A and \mathbb{C}_B of given pairs $\mathbb{R}_p = \frac{\mathbb{C}_A + \mathbb{C}_B}{2}$. \mathbb{R}_p thus, avoids erroneous choice of referencing one from a pair. As in Fig. 4, the line rotates over a cycle, and intersecting with object \mathbb{X} (A or B), generates a boolean histogram \mathcal{H} ,

$$\mathcal{H}(\mathbb{X}, \mathbb{R}_p) = [I(\mathbb{R}_p, j\Theta)]_{j=0..m-1},$$

where

$$I(\mathbb{R}_p, \theta_i) = \begin{cases} 1 & \text{if } \text{line}(\mathbb{R}_p, \theta_i) \cap \mathbb{X} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

This is extended *wlog* to the sector defined by two successive angle values (θ_i and θ_{i+1}) and is normalised with respect to the total area of the studied object such that $\sum \mathcal{H}\mathbb{X}(\cdot) = 1$. It applies to both spatial objects.

Following Fig. 4, any angle made by \mathbb{R}_p and $\mathbb{C}_\mathbb{X}$ is re-projected on the horizontal axis such as to make the histogram rotation invariant. Also, translation does not affect at all, since it uses $\mathbb{C}_\mathbb{X}$. Fig. 5 shows an example. In a similar manner, scaling does not produce any difference in \mathcal{H} as it is normalised in every sector made by θ_i and θ_{i+1} .

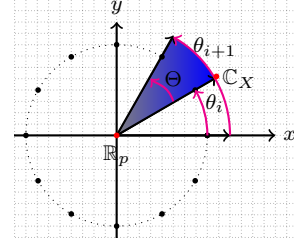


Figure 4. Computing spatial relations using radial-line rotation \odot .

B. Vertex Signature

Each vertex has a distinct vocabulary type containing different shape and size information. Since spatial relations only encode relative positioning and point distributions, and do not completely exploit global shape information, we study the pertinence of \mathcal{R} -signature [13], region based Zernike Moments [8], Generic Fourier Descriptors (GFD) [20] and Shape Context (SC) [2].

III. SYMBOL RECOGNITION

Based on our symbol description, matching of two symbols is done by matching their corresponding ARGs. Let us consider two ARGs, query $G^q = (V^q, E^q, F_A^q, F_E^q)$ and database $G^d = (V^d, E^d, F_A^d, F_E^d)$, where the set of vertices $V = \{\mathbb{T}_1, \dots, \mathbb{T}_t\}$, and the set of edges $E = \{E_1, \dots, E_r\}$.

Our matching is straightforward *i.e.*, matching has been made between the candidates only having the exact same set of vertices as well as exact labels. To generalise it, we define a binary indicator function $\tau_A^V : \Sigma_{\mathbb{T}} \rightarrow \{0, 1\}$ to check the presence of vertices in the ARG, where the value of $\tau_A^V(\mathbb{T})$ is 1 if \mathbb{T} is present in V and 0, otherwise. Now, we can then set up bijective matching functions: $\sigma : V^q \rightarrow V^d$ and $\varphi : E^q \rightarrow E^d$, respectively for vertices and edges. Now, the fusion of both alignments provides the distance between two matched graphs G^q and G^d ,

$$D(G^q, G^d) = \sum_{t \in V} \delta(F_A^q(t), F_A^d(\sigma(t))) + \sum_{r \in E} \delta(F_E^q(r), F_E^d(\varphi(r))),$$

where $\delta(a, b) = \sum_{i=1}^L \|a_i - b_i\|_2$.

IV. EXPERIMENT

A. Dataset, Ground-truth and Evaluation Metric

Fig. 1 gives some examples of very similar and very different symbols. Symbols may either be very similar in shape – and only differ by slight details – or either be completely different from a visual point of view. Symbols may also be composed of other known and significant symbols and need not necessary be connected. The global dataset is composed of roughly 500 different known symbols.

Since there is no absolute ground-truth associated to our dataset, we have asked 6 volunteers to manually select what they consider as “similar” symbols. They have chosen the candidates which have similar visual overall appearance or

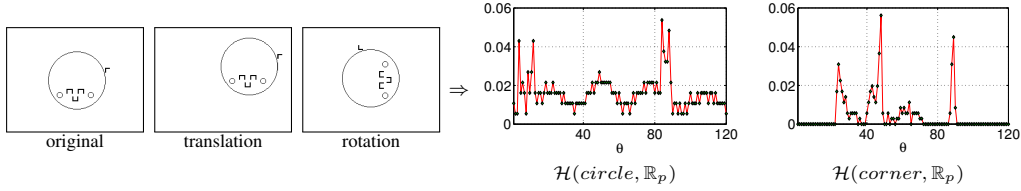


Figure 5. identical histograms at 3° resolution for original, translation and rotation between *circle* and *corner* from symbol in Fig. 2.

contain significantly similar parts with respect to the chosen query. In our testing protocol, we consider that a result returned from an algorithm is correct if at least one evaluator has selected the same result among the similar items.

For every query, we rank the symbols at the output based on distance measure as described in Section III. Since the number of ground-truths varies from one query to another, we use retrieval efficiency [6]. For a chosen query and for a given short-list of size K , it can be expressed as,

$$\eta_K = \begin{cases} n/N & \text{if } N \leq K \\ n/K & \text{otherwise,} \end{cases}$$

where n is the number of returned relevant symbols and N the total number of relevant symbols. The aim of the metric is not to miss similar candidates in the short-list since η_K computes the traditional recall if $N \leq K$ and precision measure otherwise.

B. Results and Discussions

We perform a series of tests, focussing on three major issues one-after-another as shown in Fig. 6.

Test 1. We consider the influence of different resolutions Θ in our edge signature. Its value represents the trade-off between the optimal choice of resolution – and thus precision of spatio-structural information capture – and time/space requirements. Following results in Fig. 6 (a), and given the relatively low gain of efficiency between 3° and 1° , we adopt the former for the rest of our experiments. Our spatial relation is then compared with spatial relation models: Cone-shaped [11], Angle Histogram [18] and MBR [12] as shown in Fig. 6 (b). Our edge signature outperforms them with a substantial difference.

Test 2. We employ a major set of global signal based descriptors: \mathcal{R} -signature [13], Zernike [8], GFD [20] and SC [2]. We first employ them as vertex signatures only and then confront them with the same shape descriptors, applied to the overall shape. This comparison is illustrated in Fig. 6 (c) and (d), respectively. On the whole, GFD provides interesting results in both cases. SC is restricted by the number of sample points *i.e.*, we have images ranging from a few tens of pixels to thousands of pixels.

Test 3. We integrate edge and vertex signature keeping ARG as described in Section II. While integrating, use of signatures on all vertices is not wise since our relation carries sufficient information for a few vocabulary types in a pair. Therefore, we have examined the use of signature from one

type to another as well as their possible combinations. In our tests, substantial advancement is achieved from the combination of vertices, labelled with *thick* and *circle* vocabulary types. Fig. 6 (e) shows results from using different shape signatures integrating with edge signature at 3° resolution.

In Fig. 6 (f), a comparison is made among the best of all experiments: MBR from the basic spatial relation model and GFD from shape descriptors. Our method outperforms all. For a few queries, visual illustration is shown in Fig. 7.

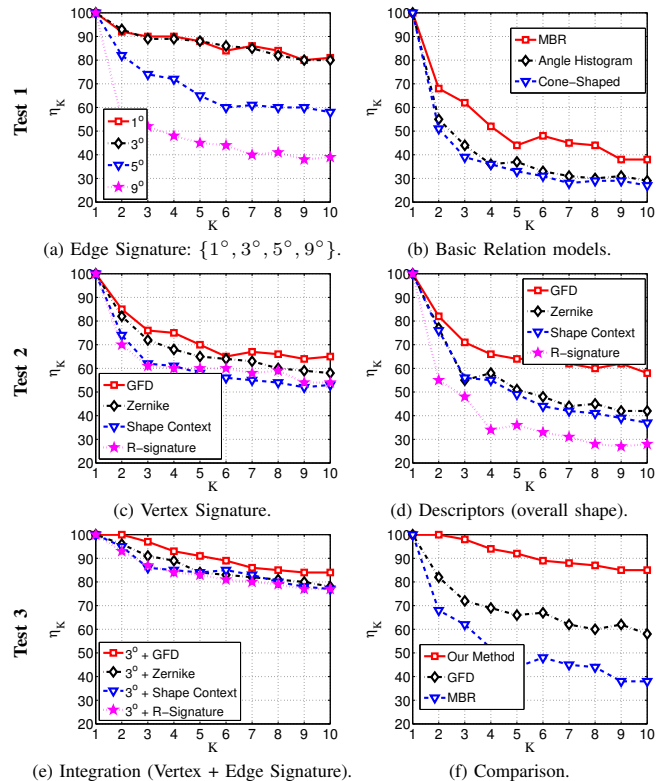


Figure 6. Average Retrieval Efficiency over 30 queries.

V. CONCLUSIONS AND FURTHER WORK

We have presented an ARG based symbol description method, where relational signatures formalise all possible connections between the vocabulary types which are labelled with global shape signatures. Our method has proven to significantly outperform state-of-the-art basic spatial relation models as well as global signal based descriptors.

	MBR			GFD			Our method											
	Q1	Q2	Q3	Q1	Q2	Q3	Edge signature			Vertex signature			Integrating both					
1.																		
2.																		
3.																		
4.																		
5.																		
6.																		
7.																		
8.																		
9.																		
10.																		

Figure 7. Visual illustration of symbol ranking at the output for a few queries: Q1, Q2 and Q3, showing ✓ for true retrieval and false, otherwise. The first symbol on the top always corresponds to the chosen query. Symbols are ranked from top to bottom based on decreasing order of similarity.

Comparison to a few more methods like [19], [22], is our next step.

REFERENCES

- [1] S. Barrat and S. Tabbone. A bayesian network for combining descriptors: application to symbol recognition. *IJDAR*, 13(1):65–75, 2010.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE PAMI*, 24(4):509–522, 2002.
- [3] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *IJPRAI*, 18(3):265–298, 2004.
- [4] L. P. Cordella and M. Vento. Symbol recognition in documents: a collection of techniques? *IJDAR*, 3(2):73–88, 2000.
- [5] J. Freeman. The modelling of spatial relations. *CGIP*, 4:156–171, 1975.
- [6] M. S. Kankanhalli, B. M. Mehtre, and J. K. Wu. Cluster-based color matching for image retrieval. *PR*, 29:701–708, 1995.
- [7] S. K.C., B. Lamiroy, and J.-P. Ropers. Inductive logic programming for symbol recognition. In *ICDAR*, pages 1330–1334, 2009.
- [8] W.-Y. Kim and Y.-S. Kim. A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication*, 16(1-2):95 – 102, 2000.
- [9] J. Lladós, E. Martí, and J. J. Villanueva. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. *IEEE PAMI*, 23(10):1137–1143, 2001.
- [10] J. Lladós, E. Valveny, G. Sánchez, and E. Martí. Symbol Recognition: Current Advances and Perspectives. In *GREC – Algorithms and Applications*, volume 2390, pages 104–127, 2002.
- [11] K. Miyajima and A. Ralescu. Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations. *Fuzzy Sets and Systems*, 2(65):225–236, 1994.
- [12] D. Papadias and Y. Theodoridis. Spatial relations, minimum bounding rectangles, and spatial data structures. *Intl. Journal of Geographical Information Science*, 11(2):111–138, 1997.
- [13] S. Tabbone, L. Wendling, and J.-P. Salmon. A new shape descriptor defined on the radon transform. *CVIU*, 102(1):42–51, 2006.
- [14] O. R. Terrades, E. Valveny, and S. Tabbone. Optimal classifier fusion in a non-bayesian probabilistic framework. *IEEE PAMI*, 31(9):1630–1644, 2009.
- [15] K. Tombre, C. Ah-Soon, P. Dosch, A. Habed, and G. Masini. Stable, robust and off-the-shelf methods for graphics recognition. *ICPR*, 1:406, 1998.
- [16] K. Tombre and B. Lamiroy. Pattern recognition methods for querying and browsing technical documentation. In *CIARP*, pages 504–518, 2008.
- [17] E. Valveny and E. Martí. A model for image generation and symbol recognition through the deformation of lineal shapes. *PRL*, 24(15):2857–2867, 2003.
- [18] X. Wang and J. Keller. Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches. *Fuzzy Sets and Systems*, 101:5–20, 1999.
- [19] S. Yang. Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor. *IEEE PAMI*, 27(2):278–281, 2005.
- [20] D. Zhang and G. Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17:825–848, 2002.
- [21] D. Zhang and G. Lu. Review of shape representation and description techniques. *PR*, 37(1):1–19, 2004.
- [22] W. Zhang, L. Wenyin, and K. Zhang. Symbol recognition with kernel density matching. *IEEE PAMI*, 28(12):2020–2024, 2006.