



# Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition

Santosh K.C., Bart Lamiroy, Laurent Wendling

► **To cite this version:**

Santosh K.C., Bart Lamiroy, Laurent Wendling. Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition. International Journal on Document Analysis and Recognition, Springer Verlag, 2013. hal-00824521

**HAL Id: hal-00824521**

**<https://hal.inria.fr/hal-00824521>**

Submitted on 22 May 2013

**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.

# Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition

K.C. Santosh<sup>†,\*</sup>, Bart Lamiroy<sup>†</sup>

<sup>†</sup>Université de Lorraine – LORIA (UMR 7503)

Campus Scientifique

BP 239 – 54506 Vandoeuvre-lès-Nancy Cedex, France

`Santosh.KC@loria.fr`, `Bart.Lamiroy@loria.fr`

Laurent Wendling<sup>‡</sup>

<sup>‡</sup>SIP – LIPADE, Université Paris Descartes (Paris V)

45, rue des Saints-Pères,

75270 Paris Cedex 06, France

`Laurent.Wendling@parisdescartes.fr`

## Abstract

This paper develops a structural symbol recognition method with integrated statistical features. It applies spatial organization descriptors to the identified shape features within a fixed visual vocabulary that compose a symbol. It builds an attributed relational graph expressing the spatial relations between those visual vocabulary elements. In order to adapt the chosen vocabulary features to multiple and possible specialized contexts, we study the pertinence of unsupervised clustering to capture significant shape variations within a vocabulary class and thus refine the discriminative power of the method. This unsupervised clustering relies on cross-validation between several different cluster indices. The resulting approach is capable of determining part of the pertinent vocabulary and significantly increases recognition results with respect to the state-of-the-art. It is experimentally validated on complex electrical wiring diagram symbols.

## Index Terms

Spatial Relations, Visual Vocabulary, Shape Descriptor, Unsupervised Clustering, Symbol Recognition.



## 1 INTRODUCTION

Graphics recognition has an extremely rich state-of-the-art literature in symbol recognition and localisation. However, most methods are targeted towards isolated line symbols, not for composed symbols connected to a complex environment [Cordella and Vento, 2000b; Lladós *et al.*, 2002]. Considering the problem of symbol localisation in real documents, composed of individual parts and constrained by spatial relations for instance, one needs to be able to extract visual parts, characterise their shape description and formalise the possible links that exist between them.

\*Corresponding author

This integration of spatial relations and shape description of the extracted visual parts is going to be the core of this paper.

Among the existing state-of-the-art, structural approaches provide powerful representations, conveying how parts are connected to each other. However, relations [Rusiñol *et al.*, 2010; Santosh *et al.*, 2011b] do not exploit shape information as shape descriptors do. On the other hand, especially when symbols are not in simple linear and isolated form, but are composed of many elements, sometimes exhibiting subtle differences, global signal-based shape descriptors cannot provide optimal retrieval performance.

In this paper, we aim to combine 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. This paper is the extension of previous work [Santosh *et al.*, 2011b; Tombre and Lamiroy, 2008; Santosh *et al.*, 2009] where we have established the significance of spatial relations for symbol recognition.

## 1.1 Related work

Studies related to the inclusion of spatial relations for symbol recognition, symbol representation and matching techniques as well as the performance analysis of several different structural approaches can be found in [Cordella and Vento, 2000b; Lladós *et al.*, 2002; Cordella and Vento, 2000a; Lladós *et al.*, 2001]. It seems to us that some of their limitations may be addressed by looking at their possible integration with statistical approaches.

In [Cordella and Vento, 2000a], shape analysis for symbol recognition has been comprehensively addressed. Their context mainly consists of isolated binary shapes. In parallel, statistical approaches like global signal-based descriptors [Yuen *et al.*, 1998; Kim and Kim, 2000; Tabbone *et al.*, 2006; Belongie *et al.*, 2002; Zhang and Lu, 2002, 2004] are usually quite fault tolerant to image distortions, since they tend to filter out small detail changes. This is unfortunately inconvenient in our context where 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. Moreover, the previously mentioned methods difficultly accommodate with connected or composite symbols and their major drawbacks are due to deformation, composition with other symbols (which, in [Yuen *et al.*, 1998] leads to unstable centroid detection, and thus errors in the ring projection) and occlusion over the boundary (leading to unstable tangents in shape context [Belongie *et al.*, 2002], for instance). In some cases, researchers have been integrating descriptors [Salmon *et al.*, 2007; Terrades *et al.*, 2008; Barrat and Tabbone, 2010], or have been combining several classifiers [Terrades *et al.*, 2009] to increase their performance, partially based on the idea presented in [Tombre *et al.*, 1998] that off-the-shelf methods are primarily designed for applications where line symbols are isolated. In these statistical approaches, signatures are simple with low computational cost. However, discrimination power and robustness strongly depend on the selection of an optimal set of features for each specific application.

Besides global shape-based symbol description, another idea is to decompose the symbols into either vector based primitives like points, lines, arcs, *etc.* or into meaningful parts like

*circles, triangles, rectangles, etc.* These methods fall under structural approaches. They are then represented as attributed relational graphs (ARG) [Bunke and Messmer, 1995; Conte *et al.*, 2004], region adjacency graphs (RAG) [Lladós *et al.*, 2001], constraint networks [Ah-Soon and Tombre, 2001] as well as deformable templates [Valveny and Martí, 2003]. Their common drawback comes from error-prone raster-to-vector conversion. Those errors can increase confusions among different symbols. Furthermore, variability of the size of graphs leads to computational complexity in matching. However, structural approaches provide a powerful representation, conveying how parts are connected to each other, while also preserving generality and extensibility. Several other approaches are more focused on computing symbol signatures by taking some regions of interest in the document image [Dosch and Lladós, 2004; Rusiñol and Lladós, 2006; Wenyin *et al.*, 2007]. These methods aim at providing faster matching in comparison to graph matching. On the other hand, they are dependent on the region of interest detector.

The overall conclusion is that one needs an appropriate image description so that the advantages of statistical features can be integrated with the expressiveness of structural approaches, thus providing generality and extensibility properties. This was already mentioned in [Tombre, 2010]:

*“... the very structural and spatial nature of the information we work with makes structural methods quite natural in the community. Their efficient integration into methods which also take full advantage of statistical learning and classification is certainly the right path to take.”*

An interesting example that uses shape descriptions and relations to form a RAG is found in [Bodic *et al.*, 2009]. The vector-based RAG is based on segmented regions which are labelled as vertices and geometric properties of adjacency relations are used to label edges. However, the approach is limited once segmented regions change with image transformations.

In the framework of stroke-based hand-drawn symbol recognition, two studies are interesting to consider [Kara and Stahovich, 2005; Lee *et al.*, 2007]. The first one is related to template-based matching. The other one uses an ARG where the vertices represent geometric primitives like lines and arcs, and the edges represent the geometric relationships between them. Matching is primarily based on graph matching or graph isomorphism [Messmer and Bunke, 2000] and is conceptually similar to [Xiaogang *et al.*, 2004]. These approaches perform well as long as the vertices are well separated and segmented (which is the case, since they are taken from on-line strokes and thus vectorisation difficulties are avoided). Recently, [Coustaty *et al.*, 2011] introduced an interesting approach where a Galois lattice is used to classify structural signatures extracted by using the Hough transform. These structural signatures are based on a topological graph using five topological relations computed between the segments as well as their lengths (based on connected and disconnected topological configurations). The paper reports that Galois lattice-based classification is very robust to noise. This however, may not provide consistent performance when symbols are found with other graphical elements or with possible texts within the image. Furthermore, since it is very dependent on the Hough transform-based segment extraction, it cannot easily incorporate more statistical shape descriptions, and therefore is not suitable for our study.

Our approach is to integrate shape descriptors with spatial relations between the visual

primitives that compose the symbol. In what follows, an outline of the proposed method will be explained.

## 1.2 Outline of the proposed method

Global signal-based descriptors can only be applied to isolated patterns. Bag-of-words approaches are widely used to overcome this problem, but require extensive training sets on the one hand, and do not take into account the global structure or arrangements between the extracted visual words. Furthermore, they usually lose the human-intuitive visual semantics of the symbol. In cases where not many training samples are available, or too costly to obtain; where the symbols' visual data itself is very redundant and overlapping, and where it is important that the symbol description matches a human sense of semantics, bag-of-words approaches are ill-suited. In previous work we addressed some of these issues [Santosh *et al.*, 2011b] and studied the possibility of having human semantics centered approaches [Santosh *et al.*, 2009].

We assume that there is no extensive training set available (in our experimental setup, we use a catalogue of known symbols for aircraft electric wiring diagrams [Tooley and Wyatt, 2008], for which only one single instance for each symbol class is available) but that there is enough human expertise available to provide knowledge of what is considered discriminant between the various symbols. Therefore, the main idea behind the approach developed in this paper is to use a set of well defined, robust, high level visual part extractors, segmenting shapes into these discriminant elementary parts, we will refer to as *visual vocabulary*. The driving motivation behind this is that a library of well mastered, robust and generic extraction tools can replace statistical bag-of-words learning techniques when insufficient learning data is available, or, as we shall show in Section 2.2, if the generic tools are not discriminant enough, it can be combined with unsupervised clustering techniques, to improve their discriminative power.

Once the symbols are segmented into their vocabulary parts, we can use the spatial relations between them [Santosh *et al.*, 2011b] to express the global spatio-structural information of the shape. However, these relations do not express shape information in the same way shape descriptors do. The spatial relation descriptors express global pixel distributions between identified areas. The areas themselves are the result of the previously mentioned vocabulary extractors.

Fig. 1 depicts how our system operates. First visual primitives are extracted from a symbol. On the one hand, they give structural information of the elementary shapes they represent. On the other hand, computing spatial relations expresses how these different shapes are positioned to one another within the symbol. The shapes are taken as nodes and the relations as arcs in an attributed relational graph representing the symbol. The recognition process relies on graph matching (which in this case is fairly trivial, as explained in Section 2, since all graph nodes are uniquely labelled and all instances of one specific vocabulary type are merged into one single node).

Unfortunately, our first experiments showed limitations when the vocabulary extraction operators were too broadly defined. The core of this paper is therefore to address this problem by introducing unsupervised clustering on the broad visual vocabulary classes, in order to refine them into more visually discriminant sub-classes. This is made possible by the fact that, once segmented from their initial images, they form a sufficient number of training samples.

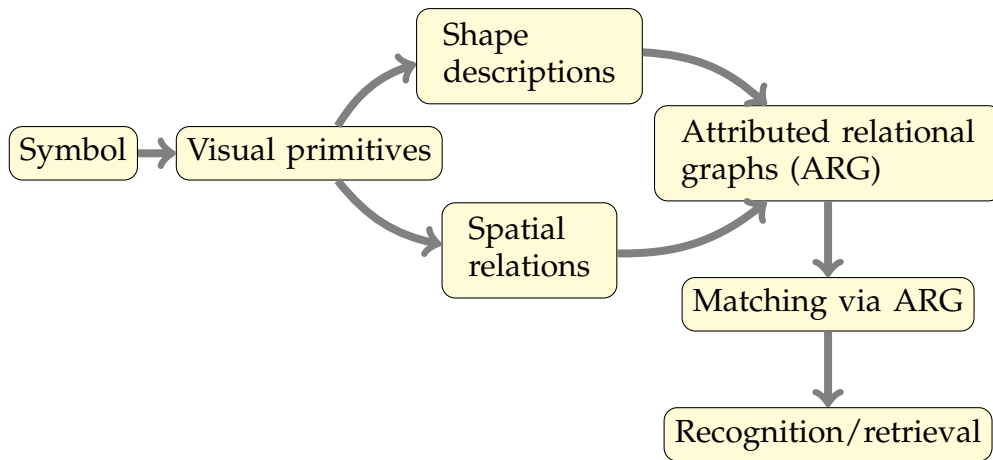


Fig. 1. An architecture of the proposed method that addresses structural method, integrated with statistical features. It uses ARG based symbol description using spatial organisation and shape features of the visual vocabulary that compose a symbol.

The remaining of the paper is organised as follows. We start by explaining our symbol description in Section 2, and more specifically how spatial relation distributions are used as arc attributes between two vertices in our ARG framework. We then discuss unsupervised vocabulary clustering via shape analysis, allowing us to sub-divide nodes into discriminant sub-shapes. In Section 3, our symbol recognition process is explained. Full experiments are reported in Section 4 and provides details of our unsupervised clustering and its validation by using several different cluster validation indices. The paper is concluded in Section 5.

## 2 SYMBOL DESCRIPTION

This section describes the general symbol description approach we use: first, we describe the visual vocabulary, how it is obtained and the resulting ARG. Second, we extend this approach by introducing a more refined analysis of the visual data segmented from the vocabulary, in order to automatically detect visually coherent sub-classes, using unsupervised classification.

### 2.1 Graph via visual vocabulary

As explained before, we are going to construct an ARG based on extracted visual components, linked together with their relative spatial distribution relations.

#### *Visual vocabulary.*

While, in the general case, the extracted vocabulary can be of any kind and from any type of features, related to what is visually pertinent in the application context under consideration, our current vocabulary is related to electrical symbols. It can be easily extended or modified by using different vocabularies and other visual cues to adapt to other domains.

We define a set of well controlled visual primitives as a *vocabulary*. They are extracted with the help of classical image analysis operators [Dosch *et al.*, 2000; Rendek *et al.*, 2004; Lamiroy

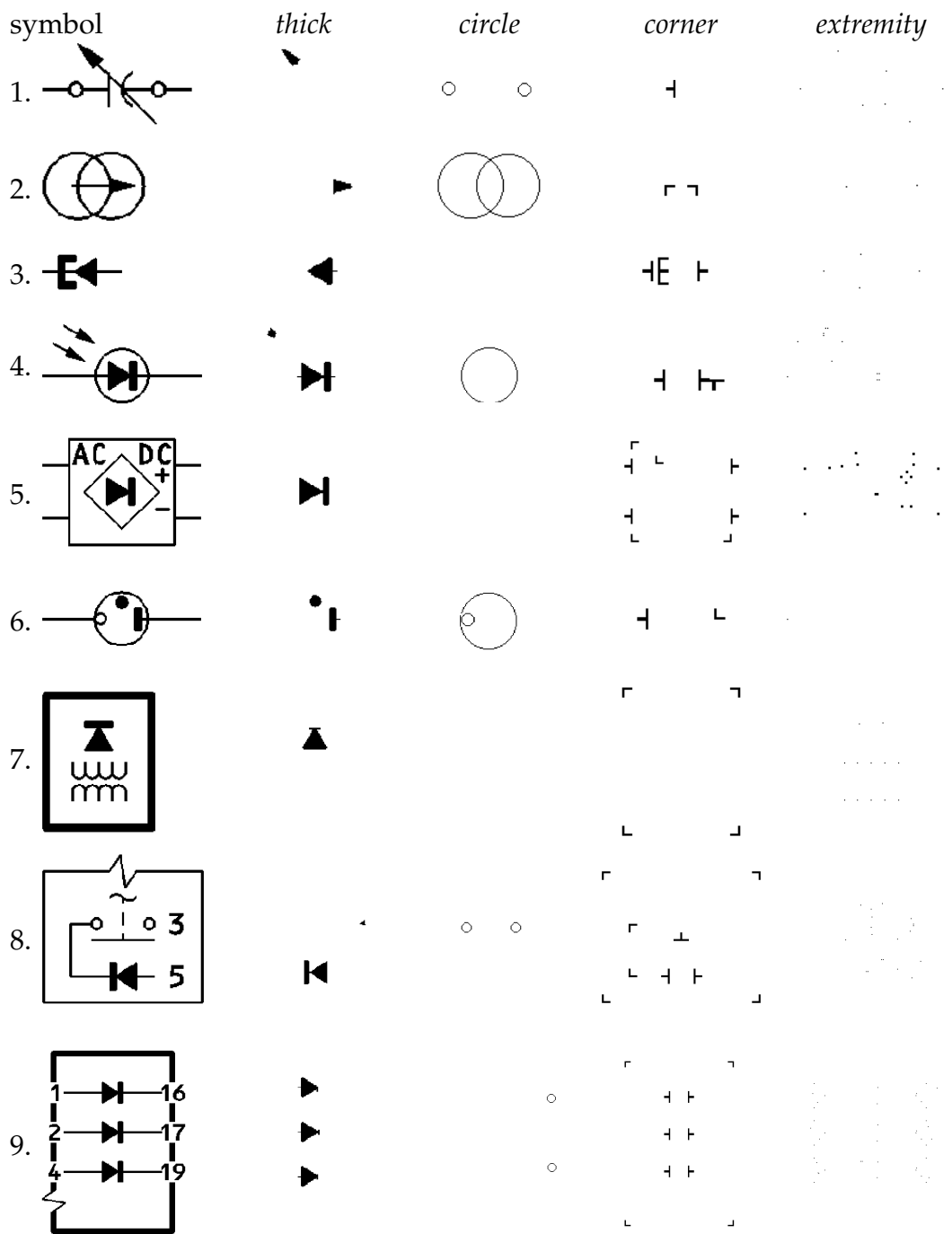


Fig. 2. Visual vocabulary from a few corresponding symbols.

and Guebbas, 2010]. Our vocabulary set consists of *circles*, *corners*, loose end *extremities* and *thick* (filled) components. Fig. 2 shows a few examples.

1) *circle* primitive

We use the algorithm as described in [Lamiroy and Guebbas, 2010] which is based on random sample consensus minimisation.

2) *corner* primitive

We only consider straight angle corners. They are extracted using a simple template match-

ing process. If the ratio of black and white pixels is greater than or equal to the template threshold, then the presence of corner is assessed.

3) *extremity* primitive

We detect loose ends from the image skeleton pixels.

4) *thick* primitive

We employ straightforward thin/thick separation by counting all *thick* connected components within the image. It takes place in a two-step process:

- a) it uses standard skeletonisation using chamfer distance and computes the histogram of line thicknesses; and
- b) an optimal cut value is computed from the histogram to distinguish between thick zones and thin zones.

On the whole, our current implementation is based on [Dosch *et al.*, 2000; Rendek *et al.*, 2004] that uses a straightforward histogram high-pass filter, following the line thickness in the document image.

In what follows, we shall refer to the set of vocabulary types as,

$$\sum_{\mathbb{T}} = \{\mathbb{T}_{thick}, \mathbb{T}_{circle}, \mathbb{T}_{corner}, \mathbb{T}_{extremity}\}. \quad (1)$$

### Graph-based representation.

Rather than using the detected elements as a basis for expressing and computing spatial relations, we group them together according to their types, as shown in Fig. 2. In a sense, a symbol is decomposed into (in our case) four layers, each of which represents either all thick components, all circles, all corners or all extremities.

The symbol can then be represented 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,
- $F_A : V \rightarrow \sum_{\mathbb{T}}$  is a function assigning attributes to the vertices. The assigned attribute is one of the extracted vocabulary classes.
- $F_E : E \rightarrow \mathcal{R}_E$  is a function assigning labels to the edges where  $\mathcal{R}$  represents spatial relations of the edge  $E$  as developed in [Santosh *et al.*, 2011b].

Effects of spatial relations on recognition performance have already been examined comprehensively for scene understanding [Biederman, 1972; Bar and Ullman, 1993], document analysis and recognition tasks [Xiaogang *et al.*, 2004; Pham and Smeulders, 2006; Rusiñol *et al.*, 2010; Santosh *et al.*, 2011b]. Based on [Freeman, 1975], spatial relations can be

- 1) topological [Egenhofer and Herring, 1991] such as *disconnected*, *externally connected*, *overlap*, and *contain or inside*;
- 2) directional [Bloch, 1999; Matsakis and Wendling, 1999; Wang and Keller, 1999] such as *north* and *south*; and
- 3) metric such as *near* or *far*.

Previously mentioned approaches address only either topological or directional relations. Managing both comes at high computational costs. Even then, no existing model fully



integrates topology. Our approach [Santosh *et al.*, 2010, 2011b] unifies both topological and directional information into one descriptor without any additional running time cost. Given two nodes,



- 1) it first uses their topological configuration to find a unique reference point;
- 2) from that reference point, a radial line is rotated over a cycle with a certain angular step, in order to cover a surrounding space; and
- 3) a directional relational histogram is computed at each angular step, accounting for the percentage of pixels of the studied visual primitives lying in that direction.

More formally, the proposed graph can be expressed as  $G = \{$

$$\begin{aligned}
 V &= \{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3, \mathbb{T}_4\}, \\
 E &= \{(\mathbb{T}_1, \mathbb{T}_2), (\mathbb{T}_1, \mathbb{T}_3), \dots, (\mathbb{T}_3, \mathbb{T}_4)\}, \\
 F_A &= \{(\mathbb{T}_1, \mathbb{T}_{thick}), (\mathbb{T}_2, \mathbb{T}_{circle}), (\mathbb{T}_3, \mathbb{T}_{corner}), (\mathbb{T}_4, \mathbb{T}_{extremity})\}, \\
 F_E &= \{((\mathbb{T}_1, \mathbb{T}_2), \mathcal{R}_{\mathbb{T}_1, \mathbb{T}_2}), ((\mathbb{T}_1, \mathbb{T}_3), \mathcal{R}_{\mathbb{T}_1, \mathbb{T}_3}), ((\mathbb{T}_2, \mathbb{T}_3), \mathcal{R}_{\mathbb{T}_2, \mathbb{T}_3})\}. \quad (2)
 \end{aligned}$$

Since this forms a complete graph, it is obvious that there exist  $r = \frac{t(t-1)}{2}$  edges for  $t$  attribute types. The major intention of having fixed and completely labelled attributes is to avoid the NP-hardness of the matching problem [Santosh *et al.*, 2011b] and also to keep coherence as vocabulary elements are semantically different.

### Limitations and graph extension.

Due to shape and size variation of the *thick* patterns, the discriminative power of this description (and therefore, the resulting retrieval performance) is sub-optimal [Santosh *et al.*, 2011b]. Taking a closer look at the extracted *thick* patterns from different symbols (Fig. 3) it is obvious that the shape and size of the *thick* pattern is related to category of the symbol from which it is assumed to be extracted. For instance, a *thick* pattern  coming from a junction is different from a triangle-shaped one, such as a significant part of a diode symbol or from an arrow: . Therefore, and in order to better distinguish these cases in the current ARG framework (*cf.* Eq. (2)), we integrate shape descriptors, and apply the shape features to label vertices.

$$F_A = \{(\mathbb{T}_1, \mathbb{T}_{thick}), \mathcal{S}_{\mathbb{T}_1}, (\mathbb{T}_2, \mathbb{T}_{circle}), (\mathbb{T}_3, \mathbb{T}_{corner}), (\mathbb{T}_4, \mathbb{T}_{extremity})\}. \quad (3)$$

Representing a single complete vertex via a global signal-based descriptor does not sufficiently exploit all available information since it is a collection of similar vocabulary type instances. We therefore introduce the following approach: since the number of *thick* patterns can be different from one symbol to another, the vertex labelled with the *thick* vocabulary type will be split into more specialised *thick* sub-vertices. This is done separately for all individual *thick* patterns.

$$\mathcal{S}_{\mathbb{T}_1} = \{s_{\mathbb{T}_1, \kappa}, \dots, s_{\mathbb{T}_1, \mathcal{K}}\}, \quad (4)$$

where  $\mathcal{K}$  is the number of *thick* patterns in a symbol. For any symbol, there thus obtain  $\{G_\kappa\}_{\kappa=1, \dots, \mathcal{K}}$  ARGs. Fig. 4 shows a description of the proposed approach.

According to our description (*cf.* Fig. 4), matching time will basically be increased in accordance with the number of *thick* patterns that compose the symbol. To reduce such a high time processing, we propose to use *thick* pattern clustering. The main idea developed in the next section is to

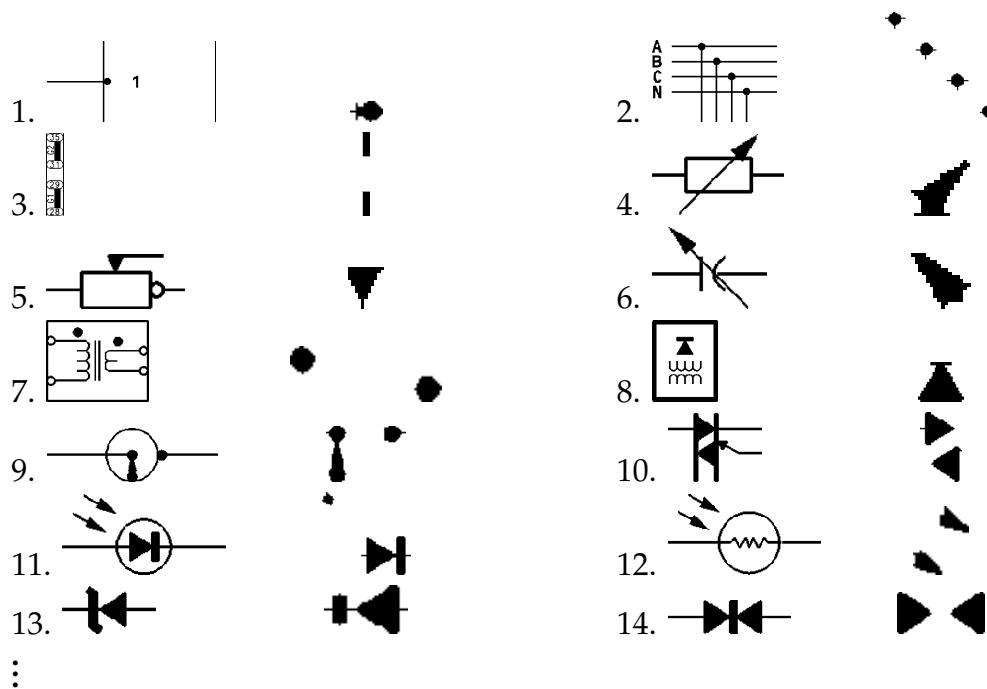


Fig. 3. Symbols and their corresponding *thick* patterns. For better understanding, extracted *thick* patterns are enlarged.

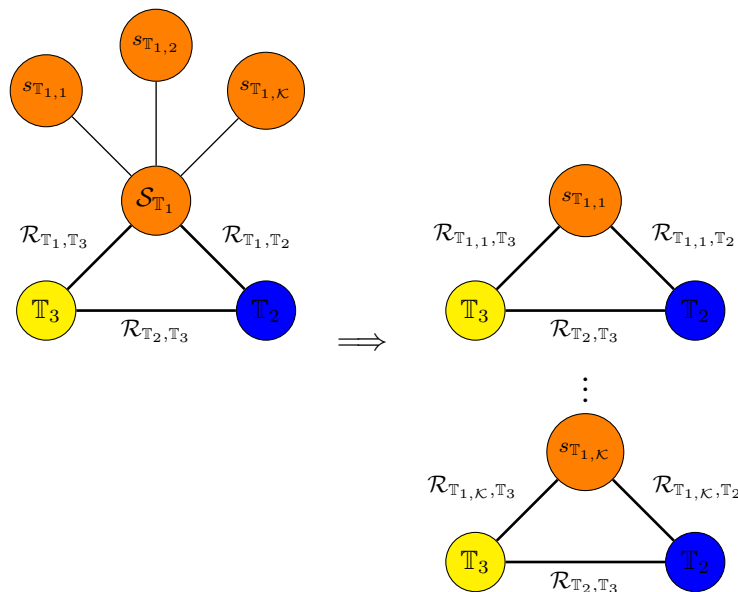


Fig. 4. A complete ARG description for a symbol with vocabulary type set *thick*, *corner* and *extremity*. The original graph is split further into a number of graphs corresponding to the number of extracted *thick* sub-classes.

refine the notion of *thick* patterns, by collectively considering all components extracted from the symbol set, to compute individual shape descriptors on each of the components, and to apply unsupervised clustering to generate a set of visually similar sub-classes within the *thick* pattern class as illustrated in Fig. 4.

## 2.2 Shape-based *thick* pattern clustering

As said before, the main idea is to perform *thick* patterns clustering using shape descriptors so that patterns with different global appearances will fall under different clusters or groups. The collected *thick* patterns in a cluster are assumed to be extracted from similar types of symbols. Since we have no *a priori* knowledge of the number of shape variations or the number of *thick* patterns in a database, we are required to perform unsupervised clustering.

### *Our clustering framework.*

We basically follow two steps [Jain and Dubes, 1988; Reynolds et al., 2006] as explained below:

**step 1.** Find the similarity or dissimilarity between every possible pair of *thick* patterns to create a distance (similarity or dissimilarity) matrix. This is achieved by representing a *thick* pattern  $p$  by a shape signature vector  $\mathbf{s}_p$  of size  $i$ . and then choosing  $\delta(\mathbf{s}_a, \mathbf{s}_b)$ , the distance (similarity or dissimilarity) metric between the signatures of *thick* patterns  $a$  and  $b$ . The outcome will of course be highly dependent on both the choice of the shape descriptor, and the metric used.

For instance, some of the more obvious metrics to construct the matrix are:

- 1) city-block:  $\delta(\mathbf{s}_a, \mathbf{s}_b) = \|\mathbf{s}_a - \mathbf{s}_b\|_1 = \sum_i |\mathbf{s}_a[i] - \mathbf{s}_b[i]|$ ,
- 2) euclidean:  $\delta(\mathbf{s}_a, \mathbf{s}_b) = \|\mathbf{s}_a - \mathbf{s}_b\|_2 = \sqrt{\sum_i (\mathbf{s}_a[i] - \mathbf{s}_b[i])^2}$  and
- 3) squared euclidean:  $\delta(\mathbf{s}_a, \mathbf{s}_b) = \|\mathbf{s}_a - \mathbf{s}_b\|_2^2 = \sum_i (\mathbf{s}_a[i] - \mathbf{s}_b[i])^2$ .

Ways to choose and experimental validation of what combination of descriptor and metric give the best results are developed further in this paper.

**step 2.** Group the similar *thick* patterns in the form of a hierarchical cluster tree.

In order to group them, we have again a choice of measuring overall distances (*linkage methods*) between clusters. We have chosen to implement three different types of linkage methods. All methods rely on the previously defined distance matrix.

- 1) Single-linkage clustering is sometimes known by *nearest neighbour* clustering. In this method, the distance between two clusters is computed as the distance between the two closest elements in two clusters. Mathematically, the distance between two clusters  $c_a$  and  $c_b$  can be expressed as,

$$D(c_a, c_b) = \min\{\delta(\mathbf{s}_a, \mathbf{s}_b) : \mathbf{s}_a \in c_a, \mathbf{s}_b \in c_b\}. \quad (5)$$

- 2) Complete-linkage clustering uses the maximum distance between the two clusters. Mathematically, it can be expressed as,

$$D(c_a, c_b) = \max\{\delta(\mathbf{s}_a, \mathbf{s}_b) : \mathbf{s}_a \in c_a, \mathbf{s}_b \in c_b\}. \quad (6)$$

- 3) Average-linkage clustering uses the mean distance between elements of each cluster.

$$D(c_a, c_b) = \frac{1}{|c_a| \times |c_b|} \sum_{\mathbf{s}_a \in c_a} \sum_{\mathbf{s}_b \in c_b} \delta(\mathbf{s}_a, \mathbf{s}_b). \quad (7)$$

Applied to our specific context, we consider the similarity matrix from all *thick* patterns. An agglomerative hierarchical clustering scheme consists in erasing rows and columns in

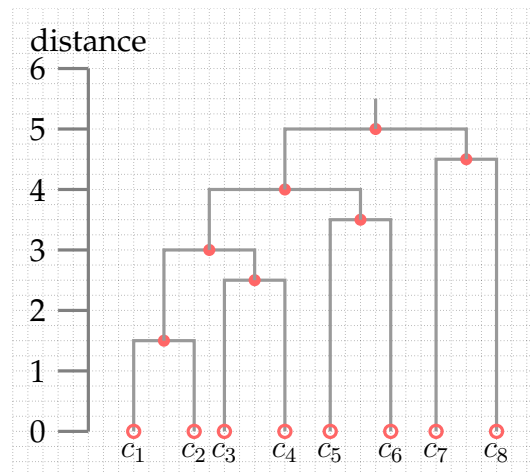


Fig. 5. Dendrogram example using eight *thick* patterns labelled with initial clusters  $c_1, c_2, \dots, c_8$ . Pairs are merged based on their similarity.

this similarity matrix each time clusters are grouped together. While grouping, similarity is based on the chosen metric and linkage method. At every merge, we update the similarity matrix by deleting the rows and columns and replacing the merged distance values by the above linkage value. We repeat the process until all clusters are merged or it reaches a pre-set cluster-threshold. Fig. 5 shows an example of a dendrogram using agglomerative hierarchical clustering. In this illustration, the clustering process ends up with a single cluster. The similarity between pairs is simply taken from one of the above linkage distance computations. For instance, clusters  $c_1$  and  $c_2$  are merged at a distance of 1.5. This is also called the *dendrogrammatic distance*.

Until now, we have simplified the whole clustering process in two steps. However, there remain two unanswered questions.

- 1) Testing all possible combinations between both descriptor metrics and linkage methods, will be time consuming. Therefore, we shall try and compute an optimal combination via cluster verification.
- 2) What is the optimal number of clusters, or how and where to cut the cluster tree? Fig. 5 only provides the clustering process until all clusters are merged, but not the cut-off threshold. To efficiently obtain the appropriate number of clusters, one either has to choose the threshold manually or set the best threshold by validating the clusters. Cluster validation requires either unsupervised or supervised approaches.

Choices and experiments related to both cluster construction and validation will be developed in the next sections.

**Cluster verification.**

Cluster analysis is highly sensitive to

- 1) shape descriptors,
- 2) distance metric selection and
- 3) linkage measure.

Approaches that use a different combination of distance metric and linkage method may yield different results (the choice of shape descriptors will be handled later). Furthermore, since the optimal combination depends on what shape signatures and data is being considered we need a dynamic approach to determine that optimum. Therefore, we are going to use the *cophenetic correlation coefficient* [Sokal and Rohlf, 1962; Rohlf and Fisher, 1968; Carr *et al.*, 1999] for every combination and choose the best one before cluster validation.

In hierarchical clustering, the height of the link represents the distance between two clusters. This height is known as cophenetic distance.

Consider the original data  $\mathcal{S} = \{\mathbf{s}_i\}$  having been clustered to produce a dendrogram  $Z$ . Let  $\bar{\delta}$  be the average value of all distance measures  $\delta(\mathbf{s}_i, \mathbf{s}_j)$  between the data samples and  $\bar{z}$  be the average of the  $Z_{i,j}$ , then the cophenetic correlation coefficient can be expressed as [Carr *et al.*, 1999]

$$\text{Cop. Cor. Coeff.} = \frac{\sum_{i < j} (\delta(\mathbf{s}_i, \mathbf{s}_j) - \bar{\delta})(Z_{i,j} - \bar{z})}{\sqrt{\left[ \sum_{i < j} (\delta(\mathbf{s}_i, \mathbf{s}_j) - \bar{\delta})^2 \right] \left[ \sum_{i < j} (Z_{i,j} - \bar{z})^2 \right]}}$$

where  $Z_{i,j}$  is the dendrogrammatic distance between the model points from the linkage function. The dendrogrammatic distance is the height of the node at which these two points are first joined together. Fig. 5 gives an example.

The cophenetic correlation coefficient expresses a combined measure between these two different sets of values: one is from distance metric and another is from linkage function. Therefore, if the clustering is valid, the linking of patterns in a cluster tree should have a strong correlation with the distance between the clusters themselves. I.e., the cophenetic correlation coefficient closer to 1 is the one corresponding to the most accurately clustered patterns. To illustrate the idea, we take a set of arbitrary features to see how the cluster verification works and how we obtain the best combination of a distance metric and a linkage method. Fig. 6 provides an example. We have provided the cophenetic correlation coefficient for all possible pairs of combinations. In this example, the combination of euclidean distance metric and the average-linkage clustering method is found to be the best compared to others because the cophenetic correlation coefficient closer to 1. This means that we do not need to use remaining pairs for cluster validation. In Fig. 6, the number of combinations can be increased by using some more distance metrics. This illustration is only intended to be a general overview of how this concept works.

### ***Cluster validation.***

In this part, we will be focussing on determining the correct number of clusters. The number of clusters has an influence on the overall aimed recognition performance. For example,

- if too many clusters are defined, they will be small in size and their elements (even inter-cluster) will be highly similar, but the analysis of many clusters increases time complexity and sensitivity to noise;

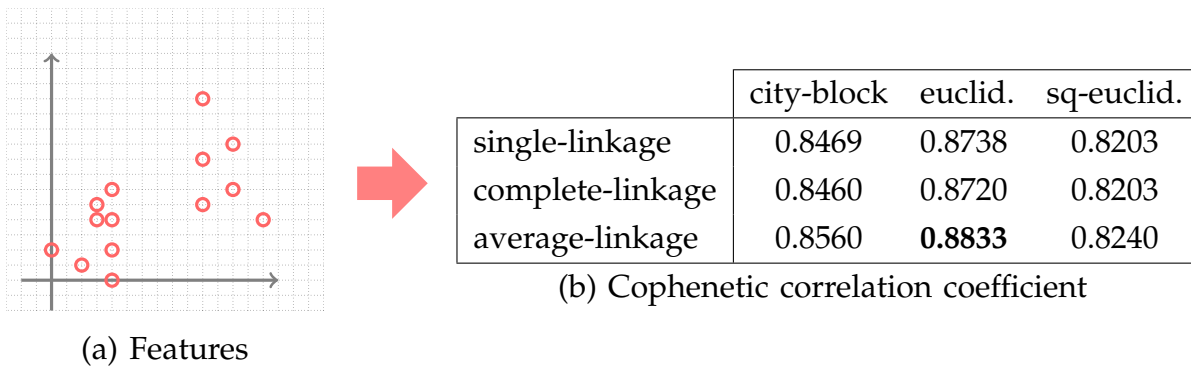


Fig. 6. Cluster verification. Cophenetic correlation coefficient from all possible combinations of distance metric and clustering linkage methods applied to a set of arbitrary features. The best combination is the one closest to 1.

- if fewer clusters are defined, they will automatically be larger, and their elements will show less similarity to one another; they will be more robust to noise, but may smooth out significant shape details.

The evaluation measures that are applied to judge various aspects of cluster validity are traditionally classified into supervised and unsupervised approaches. In our case, since we do not have external input to fix the number of clusters, we use unsupervised techniques. Unsupervised measures of cluster validity are often based on internal indices: cluster cohesion and separation.

- 1) Cluster cohesion refers to compactness or tightness of the cluster. It expresses how closely related the objects in a cluster are.
- 2) Cluster separation refers to isolation of the clusters and how distinct or well separated a cluster is from other clusters. The clusters themselves should be widely separated. There are three common approaches measuring the distance between two different clusters
  - a) the closest member of the clusters,
  - b) the most distant members, and
  - c) the centres of the clusters.

Within the framework, the indices we are going to use to validate clusters are as follows:

- 1) Dunn index [Dunn, 1974]

It is defined as the ratio between the minimal intra-cluster distance to maximal inter-cluster distance. The Dunn index (DU) for  $k$  clusters is expressed as

$$DU_k = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left( \frac{\text{dist.}(c_i, c_j)}{\max_{m=1, \dots, k} \text{dist.}(c_m)} \right) \right\}, \quad (8)$$

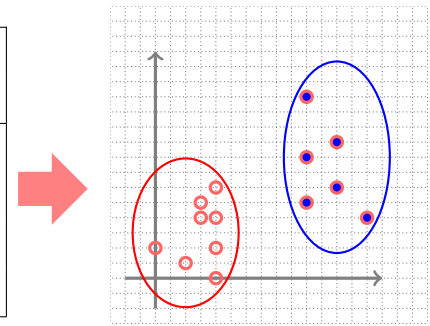
and  $\text{dist.}(c_i, c_j) = \min_{s_a \in c_i, s_b \in c_j} \delta(s_a, s_b)$  and

$$\text{dist.}(c_m) = \max_{s_a, s_b \in c} \delta(s_a, s_b).$$

It tends to be maximal when the inter-cluster distances are large and the intra-cluster distances are small.

	clusters					
	1	2	3	4	5	6
Silhouette	—	<b>0.7650</b>	0.6523	0.6689	0.7059	0.6983
Dunn	—	<b>1.2878</b>	0.6695	0.7906	0.7071	0.5590
Davies-Bouldin	—	<b>0.0359</b>	0.0984	0.1329	0.0725	0.0428
Score function	0.0000	<b>1.0000</b>	0.9000	0.8700	0.5620	0.5200

(a) Cluster validation



(b) Visual illustration

Fig. 7. Cluster validation (cf. Fig. 6). It is applied for all validation indices. For a number of clusters: {1, 2, 3, 4, 5 and 6} in (a), every index produces two clusters. Visual illustration is provided on the right side.

2) Davies-Bouldin index [Davies and Bouldin, 1979]

It identifies not only the clusters which are far from each other but also their compactness. The Davies-Bouldin (DB) index is defined as

$$DB_k = \frac{1}{k} \sum_{i=1}^k \max_{j=1, \dots, k, i \neq j} \left\{ \frac{\text{dist.}(c_i) + \text{dist.}(c_j)}{\delta(c_i, c_j)} \right\}, \quad (9)$$

$$\text{and } \text{dist.}(c_i) = \frac{1}{n_i} \sum_{s_a \in c_i} \delta(s_a, s_i^{\text{mean}}),$$

where  $n_i$  is the number of elements and  $s_i^{\text{mean}}$  the centroid of cluster  $c_i$ . The DB index is minimal for the best number of clusters.

3) Silhouette index [Rousseeuw, 1987]

Silhouette index (SI) computes the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for a total data set. The silhouette is based on the comparison of cluster tightness and separation. The average silhouette width is used to decide how good the number of selected clusters is. It is an average over all observations i.e.,

$$SI_k = \frac{1}{n} \sum_{i=1}^n \frac{(\text{dist}\hat{2}_i - \text{dist}\hat{1}_i)}{\max(\text{dist}\hat{1}_i, \text{dist}\hat{2}_i)}, \quad (10)$$

where  $n$  is the total number of elements,  $\text{dist}\hat{1}_i$  is the average distance between the element  $i$  and all other elements in its own cluster and  $\text{dist}\hat{2}_i$  is the minimum of the average distance between  $i$  and elements in other clusters. SI is maximised for the best number of clusters.

4) Score function [Saitta et al., 2007]

As in the DI and DB, score function (SF) is also based on inter-class and intra-class distance. It can be expressed as follows:

a) Between class distance ( $bcd$ )

$$bcd = \frac{\sum_{i=1}^k \delta(s_i^{mean}, s_{tot.}^{mean}) \times n_i}{n \times k} \quad (11)$$

where  $k$  is the number of clusters of size  $n$ ,  $s_i^{mean}$  the centroid of cluster  $c_i$  having  $n_i$  elements and  $s_{tot.}^{mean}$  the centroid of all clusters.

b) Within class distance ( $wcd$ )

$$wcd = \sum_{i=1}^k \left( \frac{1}{n_i} \sum_{s_a \in c_i} \delta(s_a, s_i^{mean}) \right). \quad (12)$$

Now the SF is computed as,

$$SF = 1 - \left( 1/e^{bcd-wcd} \right). \quad (13)$$

The higher the value of the SF, the more appropriate the number of clusters i.e., it maximises the  $bcd$  and minimises the  $wcd$ .

To attest the cluster validation indices, we take the example from Fig. 6 and try to deduce the best number of clusters. In Fig. 7 we can see that every cluster validation index converges at two clusters.

### 3 SYMBOL RECOGNITION

#### *Graph matching principle.*

The recognition framework principally follows the corresponding relation alignment (presented in [Santosh *et al.*, 2011b, 2013]) for matching two graphs:  $G1$  and  $G2$ , where  $G(V, E, F_A, F_E)$ . Then their distance is computed as

$$\text{Dist.}(G1, G2) = \sum_{r \in E} \delta(F1_E(r), F2_E(\sigma(r))), \quad (14)$$

where

- $\delta(\cdot)$  computes the distance between two spatial relation signatures.
  - $F1$  (resp.  $F2$ ) is the function computing the spatial relation signature of an edge.
  - $\sigma : E1 \rightarrow E2$  is the correspondence function mapping edges from one graph to the other.
- It has to be noted that, because of the unique fixed labeling of node attribute types, this mapping is immediate and exact, and there is no underlying matching difficulty.

Furthermore, as described in Section 2, a symbol  $\mathcal{S}$  has  $\mathcal{K}$  number of *thicks* and we have a set of  $\{G_\kappa\}_{\kappa=1}^{\mathcal{K}}$  ARGs representing it. Note that  $\mathcal{K}$  varies from one symbol to another. Now, we can represent a query symbol  $\mathcal{S}^q = \{G_\kappa^q\}_{\kappa=1}^{\mathcal{K}}$  and a database symbol  $\mathcal{S}^{db} = \{G_{\kappa'}^{db}\}_{\kappa'=1}^{\mathcal{K}'}$ . To compute the similarity between the two symbols, the main idea is to find the best matched graphs pair. This means that we compute the distances between all possible pairs of graphs. I.e., any database graph  $G_\kappa^{db}$  that is matched with any query graph  $G_\kappa^q$  is sufficient to find similarity between the symbols:  $\mathcal{S}^q$  and  $\mathcal{S}^{db}$ .



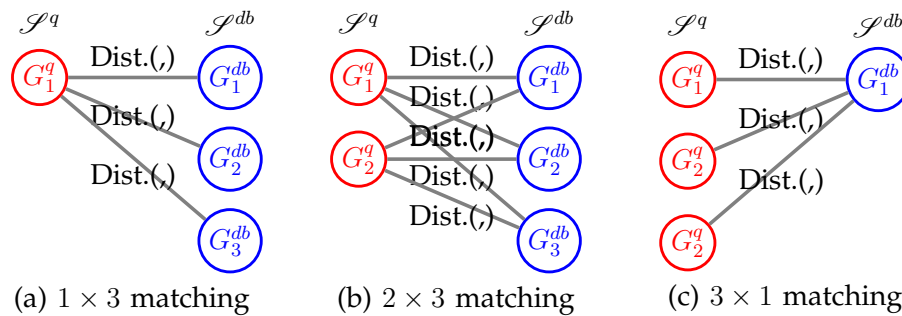


Fig. 8. A few examples showing possible graph matching scheme between two symbols  $\mathcal{S}^q$  and  $\mathcal{S}^{db}$  by providing various graph(s) that compose a symbol. In this illustration, each encircled token represents a graph. This illustrates a basic idea without integrating *thick* pattern clustering.

Let us take any pair of symbols  $\mathcal{S}^\dagger$  and  $\mathcal{S}^\ddagger$  to formally compute the minimum distance between them,

$$\Delta(\mathcal{S}^\dagger, \mathcal{S}^\ddagger) = \min_{\kappa} \left( \min_{\kappa'} \left( \text{Dist.}(G_{\kappa}^\dagger, G_{\kappa'}^\ddagger) \right) \right). \quad (15)$$

Fig. 8 shows three possible cases to realise graph matching between them, without considering *thick* pattern clustering. Considering whole database symbols  $\{\mathcal{S}^{db}\}_{db=1, \dots, \mathbb{DB}}$ , the closest candidate for any query symbol  $\mathcal{S}^q$ , can be computed as,  $\min_{db} (\Delta(\mathcal{S}^q, \mathcal{S}^{db}))$ . Note that, we are not just looking for the closest candidate, but also retrieving database symbols for any provided short-list. In the following part, we explain about how retrieval can be handled.

### *An inclusion of thick pattern clustering in graph matching.*

The aforementioned mentioned matching concept suffers from heavy computation as soon as the number of *thick* increases. To recover this, we are going to use *thick* pattern clustering, as mentioned before.

**step 1.** For each query symbol, the first step is to allocate the cluster in which the query *thick* belongs to. To compute the distance between any test *thick* pattern labelled with a shape signature, we use a centroid of the particular cluster i.e.,  $\delta(s_a, s_i^{mean})$ . It is shown in Fig. 9. More than one cluster can be selected if a query symbol has two or more *thick* patterns with different shape information. Fig. 10 shows a couple of examples of it.

**step 2.** Once the cluster(s) is(are) selected, then the symbols related to those *thick* patterns i.e., corresponding symbols are taken for matching. Our graph matching is explained in Eq. (14).

As a consequence, for matching, we have not taken

- 1) all database symbols; and
- 2) all graphs related to a particular database symbol.

This means that a limited number of database symbols are used i.e.,  $\{\mathcal{S}^{db}\}_{db=1, \dots, \widehat{\mathbb{DB}}}$ , where  $\widehat{\mathbb{DB}} \subseteq \mathbb{DB}$ . In addition, for each particular database symbol  $\mathcal{S}^{db}$ , the number of graph matchings can be reduced to  $\widehat{\mathcal{K}}' \subseteq \mathcal{K}'$ . Clustering thus, helps to reduce running time.

### *Retrieval and ranking.*

The previously defined distance  $\Delta(\cdot)$  provides an idea of how similar/dissimilar a database

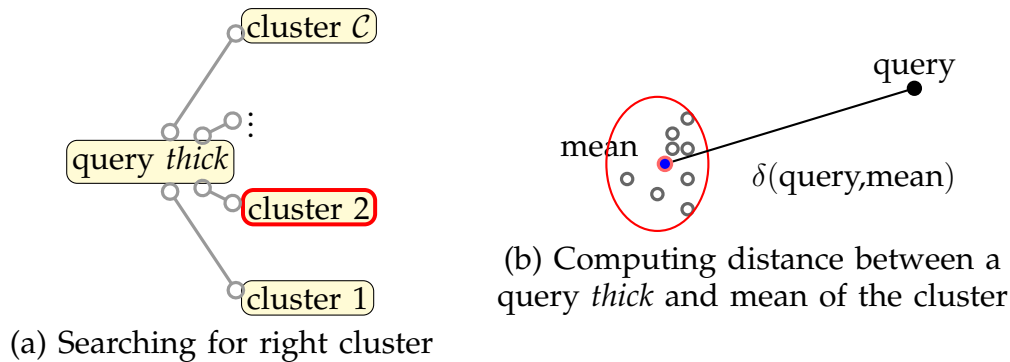


Fig. 9. Searching for the right cluster(s) where query *thick* pattern belongs to. In this illustration, the red-boxed cluster in (a) is the corresponding cluster for a query *thick*. To compute distance, only representative (mean) of every cluster will be taken into account as shown in (b).

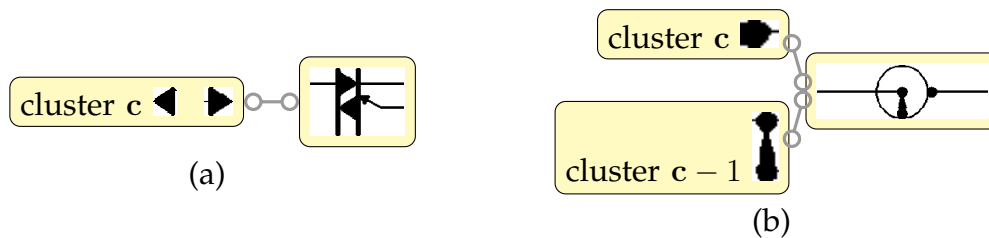


Fig. 10. *Thick* pattern(s) in cluster(s) and its (their) corresponding symbol. A symbol can have more than one cluster. It is primarily due to different shaped *thick* patterns. In (a), both *thick* patterns appear in a single cluster while there exist two different clusters in case (b).

symbol is, with respect to a query. In order for this similarity measure to fall into an appropriate range, we normalise  $\Delta(\cdot)$  to  $[0, 1]$  by taking all filtered database symbols  $\widehat{\mathbb{D}\mathbb{B}}$ ,

$$\bar{\Delta}(\cdot) = \frac{\Delta(\cdot) - \Delta(\cdot)^{min.}}{\Delta(\cdot)^{max.} - \Delta(\cdot)^{min.}} \quad (16)$$

Now the matching score,  $m.score(\mathcal{S}^q, \mathcal{S}^{db}) = 1 - \bar{\Delta}(\mathcal{S}^q, \mathcal{S}^{db})$ . For any provided short-list, ranking can be made based on the decreasing order of matching score.

## 4 EXPERIMENTS

In order to measure the impact of the vocabulary clustering compared to using relations only [Santosh *et al.*, 2011b], we have used the same dataset and experimental protocol.

We use well-known state-of-the-art of shape descriptors used for *thick* pattern (vocabulary) clustering:

- 1) Zernike moments (ZM) [Kim and Kim, 2000],
- 2)  $\mathcal{R}$ -signature [Tabbone *et al.*, 2006],
- 3) shape context (SC) [Belongie *et al.*, 2002],
- 4) generic fourier descriptors (GFD) [Zhang and Lu, 2002] and
- 5) DTW-Radon (lets say  $\mathcal{D}$ -Radon) [Santosh *et al.*, 2011a].

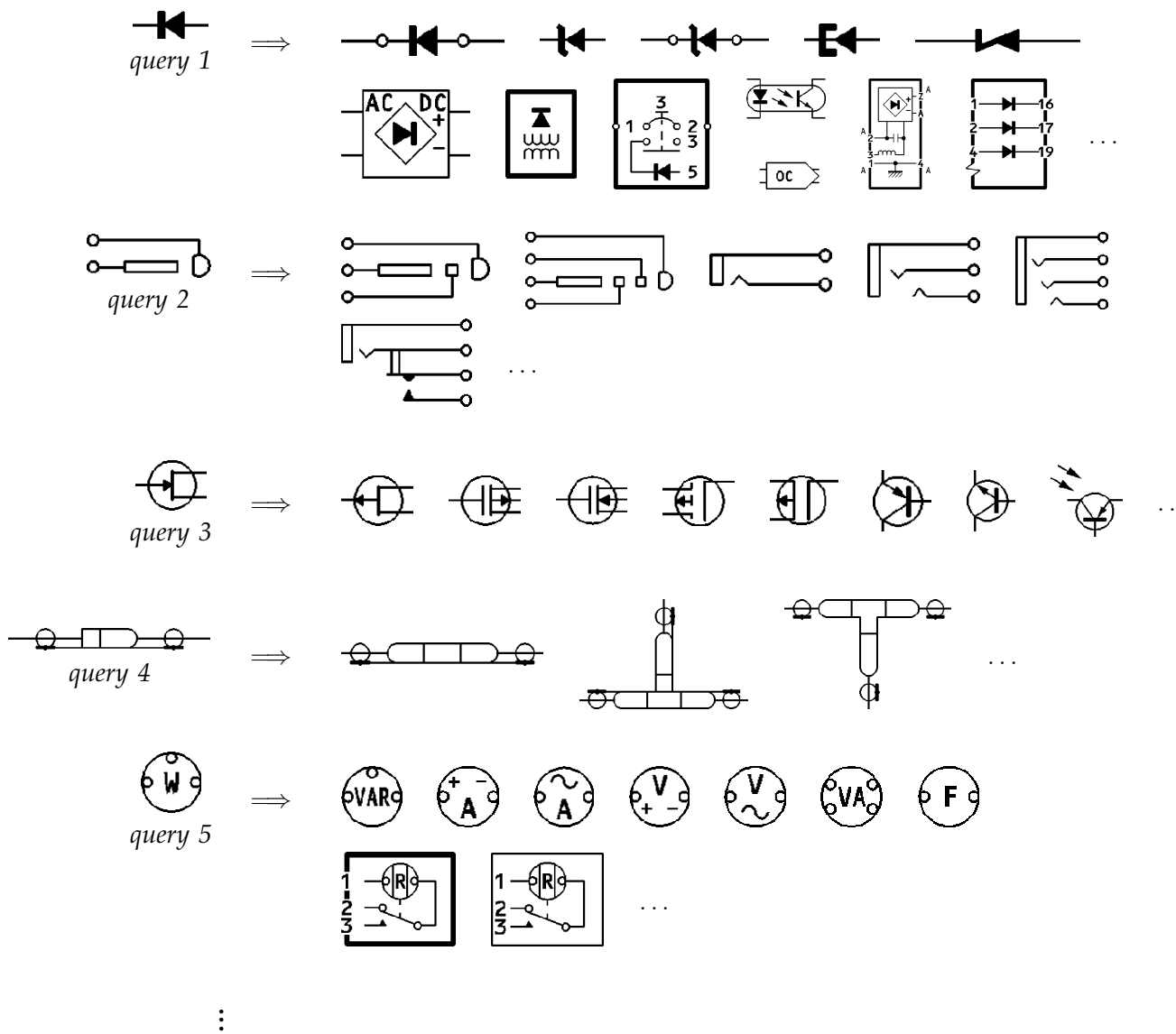


Fig. 11. A sample of electrical symbols. For every query symbol: *query 1* to *query 5*, a few human-selected relevant symbols are listed.

### 4.1 Dataset, ground-truth and evaluation metric

#### *Dataset.*

Fig. 11 gives an overview of the dataset we are using for our experiments. The global dataset is composed of roughly 500 different symbols, taken from electrical wiring diagrams [Tooley and Wyatt, 2008]. It contains symbols that are either very similar in shape – and only differ by slight details – or, on the other hand, are 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.

#### *Ground-truth and evaluation metric.*

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, and we have merged their inputs

to reduce possible subjective bias. They have chosen the candidates which have similar visual overall appearance or 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 similar items. In more formal terms, for each query, the “ground-truth” is considered to be the set of symbols formed by the union of all human selected sets. In Fig. 11, we have provided a few samples of query symbols and the corresponding lists of relevant symbols.

For every query, we rank the symbols at the output based on the distance measure described in Section 3. Since the number of similar symbols, according to the ground-truth, may vary a lot from one query to another, we use retrieval efficiency [Kankanhalli *et al.*, 1996] as a measure for retrieval quality. For a chosen query and for a fixed number of  $K$  returned symbols, it can be expressed as,

$$\eta_K = \begin{cases} n/N & \text{if } N \leq K \\ n/K & \text{otherwise,} \end{cases} \quad (17)$$

where  $n$  is the number of returned relevant symbols and  $N$  the total number of relevant symbols in the dataset. Note that  $\eta_K$  computes the traditional recall if  $N \leq K$  and computes precision otherwise. The main advantage of this is that the average retrieval efficiency curve is not biased even with different quantities of ground-truth for different queries, which happens when using the precision metric with  $N < K$ .

## 4.2 Thick pattern clustering

The goal of this section is to illustrate how choices of descriptors and cluster validation indices can have an influence on the obtained clusters. The way of actually selecting the best clustering parameters will be discussed in Section 4.3. We are considering the application of various shape signatures on *thick* patterns. Since discrimination power of shape descriptors vary from one to another, the number of clusters will obviously be different. Furthermore, a change in cluster validation indices also impacts the results, as we shall show here. We are going to test different shape descriptors and perform various clustering and cluster validation indices on each of them in Section 4.3. The following section uses GFD [Zhang and Lu, 2002] as an example.

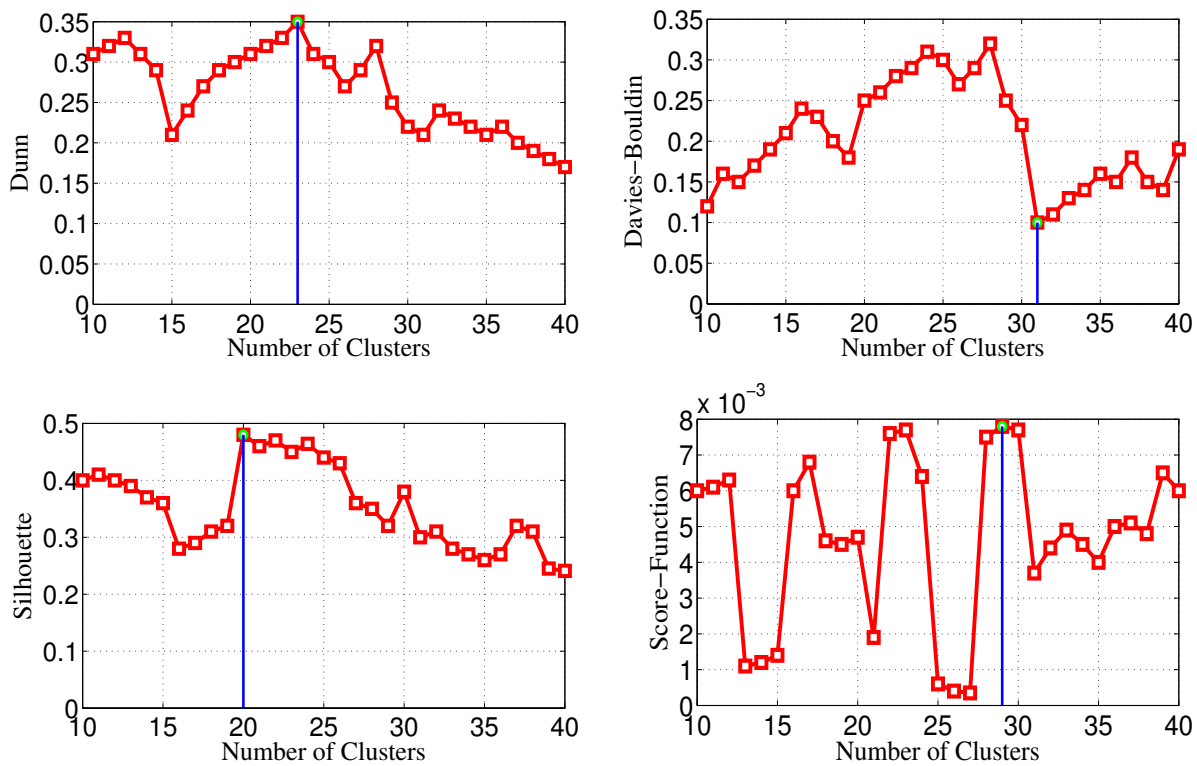
		Angular frequency							
		6	8	10	12	14	16	18	20
Radial frequency	4	7	9	10	10	16	19	21	21
	6	9	10	17	24	24	<b>33</b>	28	35
	8	9	9	21	23	33	38	38	40
	10	11	11	21	28	31	40	37	34
	12	21	27	31	33	38	33	30	28

(a) Dunn index

		Angular frequency							
		6	8	10	12	14	16	18	20
Radial frequency	4	9	12	11	12	19	21	22	24
	6	10	13	19	25	22	<b>30</b>	29	33
	8	12	14	23	25	31	34	40	40
	10	9	11	21	25	32	38	32	35
	12	21	24	27	36	36	35	33	23

(b) Davies-Bouldin index

Fig. 12. Number of clusters after cluster validation tests using various GFD parameters using Dunn and Davies-Bouldin indices.



GFD parameters: radial frequency = 6 and angular frequency = 15.

Fig. 13. An example of how cluster validation indices behave for GFD. The maximal value is shown by the vertical line in each plot.

### Clustering using GFD.

We have sampled the GFD parameters from 4 to 12 for the radial, and from 6 to 20 for angular frequencies. For each of these values, we have computed the Dunn and Davies-Bouldin clustering indices. Fig. 12 shows the effect of these changes on the number of clusters obtained. For example, we get 33 clusters from Dunn and 30 clusters from Davies-Bouldin when we take radial and angular frequencies values of (6,16). For more clarity, each value in Fig. 12 (i.e., number of clusters in red box, for instance), we have provided a complete illustration in Fig. 13 about how cluster validation indices select different number of clusters. The observation is that depending on the chosen parameters (i.e., GFD parameters) and criteria (i.e., validation indices), the final number of clusters can vary widely.

Using this framework, and for visual illustration, Fig. 14 shows a few clusters and their corresponding patterns from the Davies-Bouldin cluster validation index.

### Clustering using other shape descriptors.

For the remaining shape descriptors, mentioned before, like shape context, Zernike moments,  $\mathcal{R}$ -signature,  $\mathcal{D}$ -Radon, a similar cluster validation process has been carried out. In Section 4.3, the effect of the use of all shape descriptors on retrieval performance will be presented. This goes with the appropriate choice of cluster validation index.

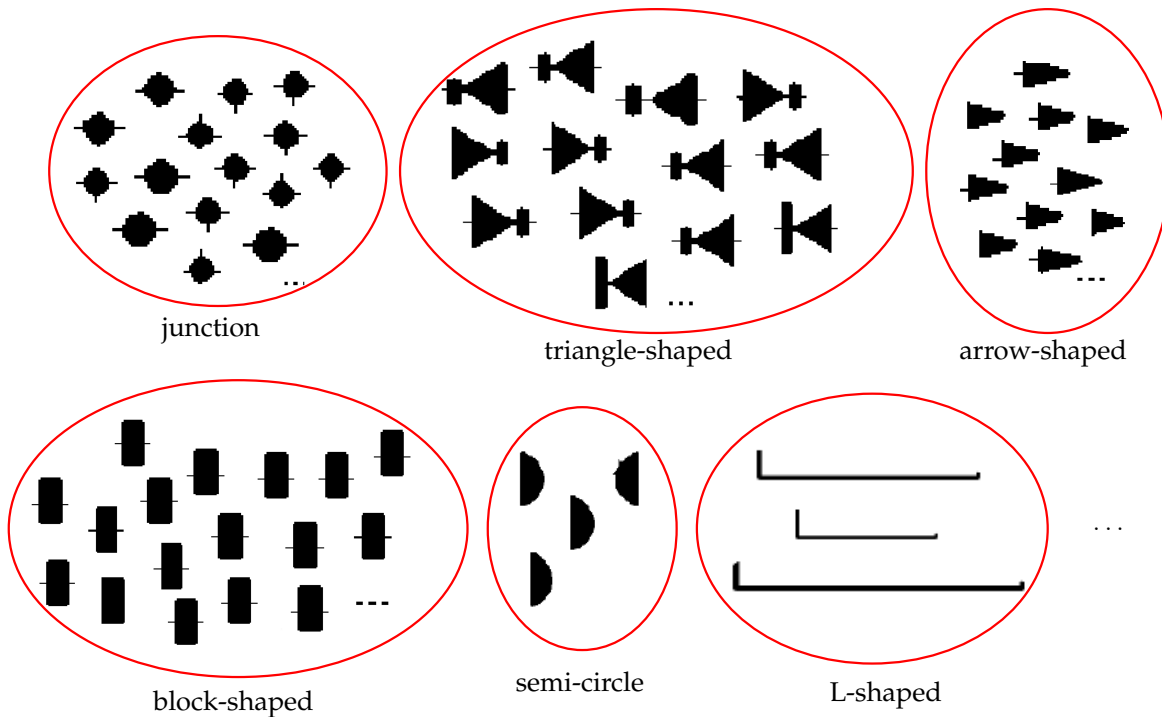


Fig. 14. A selection of pattern clustering results on *Thick* components, using GFD shape descriptor and Davies-Bouldin.

### 4.3 Results and analysis

#### *Related previous work.*

Since the approach is the extension of a previous work, we admit some of the points and conclusions described in [Santosh *et al.*, 2011b] and get directly to the point for comparisons with the existing state-of-the-art. For clarity, we recall some of the experimental setup choices, such as the use of retrieval efficiency, as described in Section 4.1, which has been systematically computed for values of  $K = 1$  to 10, over 30 queries, for all experiments described in this section.

- 1) In the previous approach, called RLM [Santosh *et al.*, 2011b], we have studied the influence of different resolutions for the angular sampling steps. Its value represents the trade-off between the optimal choice of resolution – and thus precision of spatio-structural information capture – and time/space requirements. As reported in [Santosh *et al.*, 2011b], the best trade-off is  $3^\circ$ , which is the value we adopt for the rest of our experiments. During its comparison with most state-of-the-art spatial relation models such as cone-shaped [Miyajima and Ralescu, 1994], angle histogram [Wang and Keller, 1999] and MBR [Papadias and Theodoridis, 1997], MBR came out as the the best performer among the three competing benchmarking relation models. RLM performed better than MBR with a substantial difference of more than 30%.
- 2) Similarly, a series of experiments reported in [Santosh *et al.*, 2011a,b] study the appropriateness of a set of major state-of-the-art shape descriptors such as Zernike moments (ZM) [Kim and Kim, 2000],  $\mathcal{R}$ -signature [Tabbone *et al.*, 2006], shape context (SC) [Belongie *et al.*, 2002], generic Fourier descriptors (GFD) [Zhang and Lu, 2002] and  $\mathcal{D}$ -Radon [Santosh

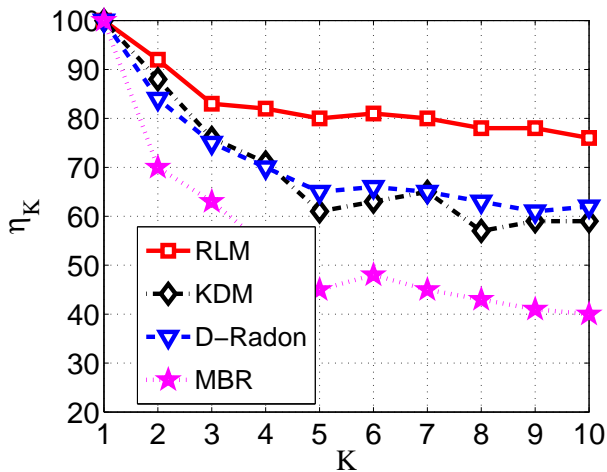


Fig. 15. Average retrieval efficiency with  $K = 1$  to 10 over 30 queries.

*et al.*, 2011a].  $\mathcal{D}$ -Radon performs best, but with a 12% lower performance, on average, with respect to RLM.

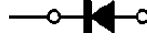

- Furthermore, we have also compared RLM to pixel-based approaches specially designed for symbol recognition: statistical integration of histogram array (SIHA) [Yang, 2005] and Kernel density matching (KDM) [Zhang *et al.*, 2006], using the same protocol as described before. In this comparison, RLM outperforms KDM, the best performer of the two tested descriptors.

Fig. 15 gives an overview of the observed performance differences between all tested methods.

### *Integrating vocabulary clustering with spatial relations.*

Following the reported results in Fig. 15, in this section, we aim to see how the integration of *thick* pattern selection (with spatial relations) improves the retrieval performance. Our query *thick* pattern selection is based on clustering mentioned in Section 2.2. Clustering performance is based on shape signatures and cluster validation indices. Therefore, we take both into account in order to assess a suitable combination for this application.

Fig. 16 shows the comparison of performance of cluster validation indices for different shape descriptors. In the tests, we observe and analyse retrieval performances on a one-to-one basis. Overall, GFD outperforms all, but  $\mathcal{D}$ -Radon performs almost equally having a marginal difference. Zernike moments, shape context and  $\mathcal{R}$ -signature are lagging behind.

Selection of shape descriptors does not only provide a complete process while it needs to account cluster validation indices, on the other hand. As said previously, different cluster validation indices provide different results. As a consequence, overall retrieval performance is affected. As shown in Fig. 16, for all shape descriptions, Dunn and Davies-Bouldin indices provide almost similar advancements, while the remaining silhouette and sore-function do not. Therefore, either Dunn or Davies-Bouldin index will be the appropriate choice for this application. For a quick visual illustration, as an example  is taken as a query symbol. Using  $\mathcal{D}$ -Radon to describe the shape of *thick* , the first seven ranked database symbols retrieved are as follows,

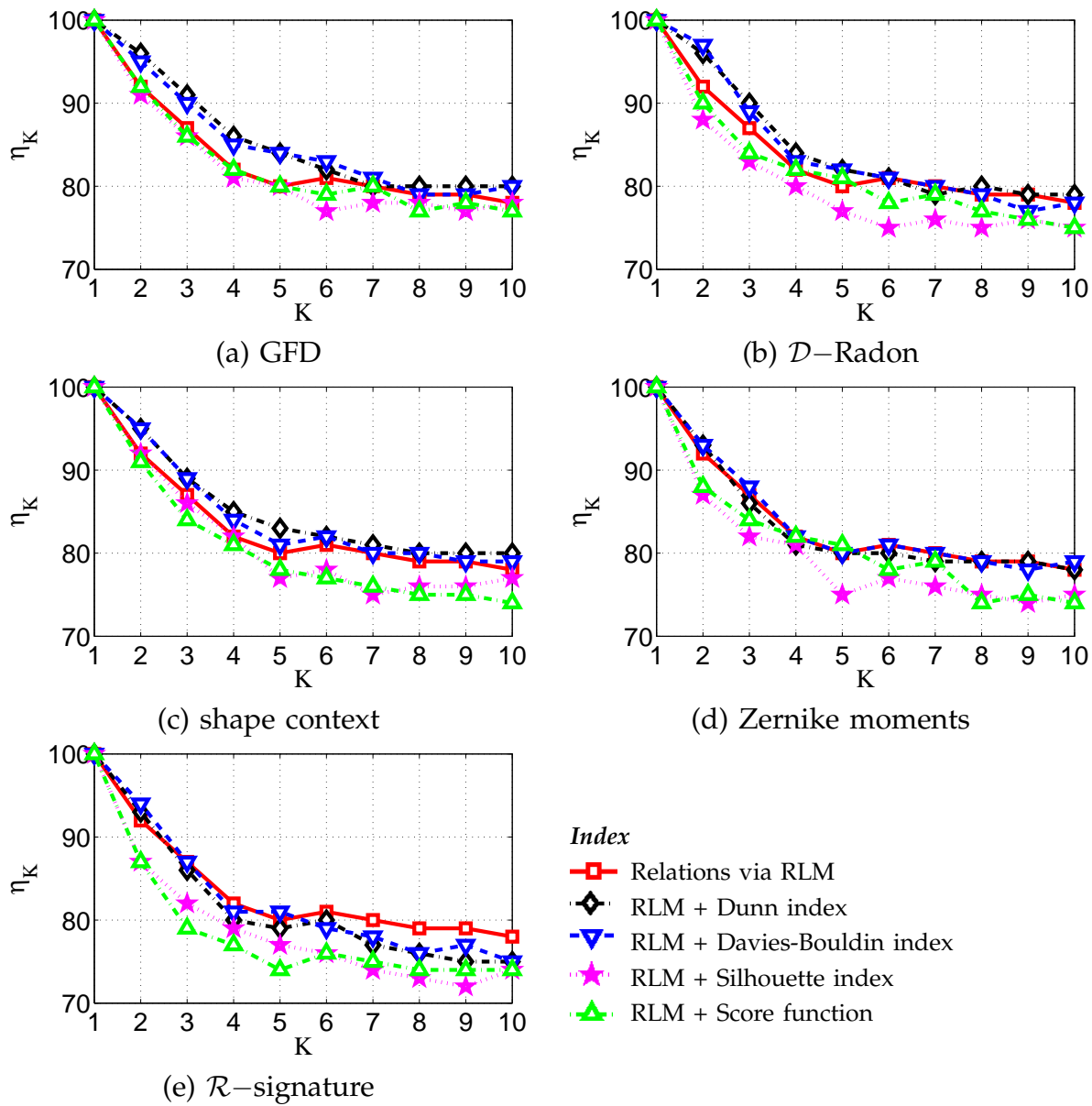
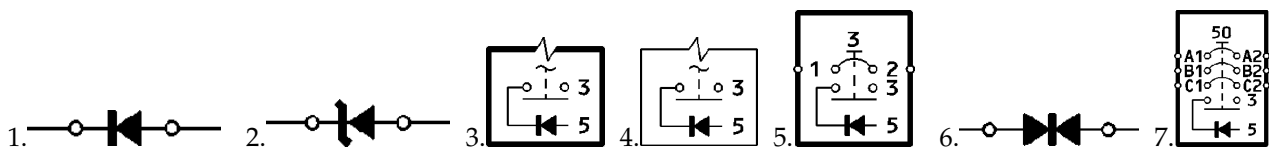


Fig. 16. Average retrieval efficiency over requested list: 1 to 10 using global signal-based shape descriptors for *thick* patterns clustering and several different cluster validation indices.



A few more queries are illustrated in Fig. 17 where  $\mathcal{D}$ -Radon is used to describe the shape of the *thick* for clustering and Davies-Bouldin index is used to validate the cluster. In this illustration, the first symbol on the top always correspond to the query symbol and retrieved symbols are ranked from top to bottom based on the order of similarity. These retrieval examples provide an idea of how far the retrieval performance after integrating vocabulary clustering, has been advanced. After integrating vocabulary clustering, we have found that retrieval symbols are visually better ranked in addition to its efficacy.



Ranking ↓	Relation			Vocabulary clustering + Relation		
	Q1	Q2	Q3	Q1	Q2	Q3
1.						
2.						
3.						
4.						
5.						
6.			CO			
7.			HP			
8.						
9.						
10.						

Fig. 17. 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. It uses  $\mathcal{D}$ -Radon to describe shape of the *thick* pattern in clustering and Davies-Bouldin index to validate cluster.

Along with relational signature matching, query *thick* pattern selection via clustering advances retrieval performance. However, no surprising difference is observed. The major reason is that not all query symbols contain *thick* pattern in their vocabulary type sets. In other words, absence of *thick* vocabulary type means ranking has been made only through relation alignment, with eventually no change in retrieval performance.

## 5 CONCLUSIONS

In this paper, we have presented the concept to describe the shape of the extracted visual vocabulary that shows significant shape variations, and integrate relations that exist between them. Keeping the ARG framework, our intelligent concept of using shape descriptor via clustering provides an immediate retrieval applications. We have comprehensively studied unsupervised clustering and evaluated with several different well-known validation indices. In our tests, we have observed the behaviour of the well-known state-of-the-art shape descriptors.

Very specifically, we have employed *thick* pattern clustering and use the related clusters with respect to the query for relation alignment. Clustering of thick patterns thus opens a global concept that it can be applied for any other visual primitives. Overall, we bring an attention to the use of a hybrid approach in symbol recognition since it combines both worlds: structural and statistical.

## REFERENCES

- Ah-Soon, C. and Tombre, K. (2001). Architectural symbol recognition using a network of constraints. *Pattern Recognition Letters*, **22**(2), 231–248.
- Bar, M. and Ullman, S. (1993). Spatial context in recognition. *Perception*, **25**, 324–352.
- Barrat, S. and Tabbone, S. (2010). A bayesian network for combining descriptors: application to symbol recognition. *International Journal on Document Analysis and Recognition*, **13**(1), 65–75.
- Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(4), 509–522.
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, **177**(43), 77–80.
- Bloch, I. (1999). Fuzzy relative position between objects in image processing: New definition and properties based on a morphological approach. *Uncertainty Fuzziness and Knowledge-Based Systems*, **7**(2), 99–133.
- Bodic, P. L., Locteau, H., Adam, S., Héroux, P., Lecourtier, Y., and Knippel, A. (2009). Symbol detection using region adjacency graphs and integer linear programming. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 1320–1324.
- Bunke, H. and Messmer, B. T. (1995). Efficient attributed graph matching and its application to image analysis. In C. Braccini, L. D. Floriani, and G. Vernazza, editors, *Proceedings of International Conference on Image Analysis and Processing*, volume 974 of *Lecture Notes in Computer Science*, pages 45–55. Springer-Verlag.
- Carr, D. B., Young, C. J., Aster, R. C., and Zhang, X. (1999). Cluster analysis for ctb seismic event monitoring. In *A study prepared for the U.S. Department of Energy*.
- Conte, D., Foggia, P., Sansone, C., and Vento, M. (2004). Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, **18**(3), 265–298.
- Cordella, L. P. and Vento, M. (2000a). Symbol and shape recognition. In A. K. Chhabra and D. Dori, editors, *Graphics Recognition, Recent Advances*, volume 1941 of *Lecture Notes in Computer Science*, pages 167–182. Springer.
- Cordella, L. P. and Vento, M. (2000b). Symbol recognition in documents: a collection of techniques? *International Journal on Document Analysis and Recognition*, **3**(2), 73–88.
- Coustaty, M., Bertet, K., Visani, M., and Ogier, J.-M. (2011). A new adaptive structural signature for symbol recognition by using a galois lattice as a classifier. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, **41**(4), 1136–1148.
- Davies, D. L. and Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-1**(2), 224–227.
- Dosch, P. and Lladós, J. (2004). Vectorial signatures for symbol discrimination. In J. Lladós

- and Y.-B. Kwon, editors, *Graphics Recognition, Recent Advances and Perspectives*, volume 3088 of *Lecture Notes in Computer Science*, pages 154–165. Springer-Verlag.
- Dosch, P., Tombre, K., Ah-Soon, C., and Masini, G. (2000). A complete system for the analysis of architectural drawings. *International Journal on Document Analysis and Recognition*, **3**(2), 102–116.
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, **4**(1), 95–104.
- Egenhofer, M. and Herring, J. R. (1991). *Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases*. University of Maine, Research Report.
- Freeman, J. (1975). The modelling of spatial relations. *Computer Graphics and Image Processing*, **4**, 156–171.
- Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Kankanhalli, M. S., Mehtre, B. M., and Wu, R. K. (1996). Cluster-based color matching for image retrieval. *Pattern Recognition*, **29**(4), 701–708.
- Kara, L. B. and Stahovich, T. F. (2005). An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers & Graphics*, **29**(4), 501–517.
- Kim, W.-Y. and Kim, Y.-S. (2000). A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication*, **16**(1-2), 95 – 102.
- Lamiroy, B. and Guebbas, Y. (2010). Robust and precise circular arc detection. In J.-M. Ogier, W. Liu, and J. Lladós, editors, *Graphics Recognition. Achievements, Challenges, and Evolution, 8th International Workshop, GREC 2009, La Rochelle, France, July 22-23, 2009. Selected Papers*, volume 6020 of *Lecture Notes in Computer Science*, pages 49–60. Springer-Verlag.
- Lee, W., Kara, L. B., and Stahovich, T. F. (2007). An efficient graph-based recognizer for hand-drawn symbols. *Computers & Graphics*, **31**(4), 554–567.
- Lladós, J., Martí, E., and Villanueva, J. J. (2001). Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(10), 1137–1143.
- Lladós, J., Valveny, E., Sánchez, G., and Martí, E. (2002). Symbol Recognition: Current Advances and Perspectives. In D. Blostein and Y.-B. Kwon, editors, *GREC – Algorithms and Applications*, volume 2390 of *Lecture Notes in Computer Science*, pages 104–127. Springer-Verlag.
- Matsakis, P. and Wendling, L. (1999). A New Way to Represent the Relative Position Between Areal Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(7), 634–643.
- Messmer, B. T. and Bunke, H. (2000). Efficient subgraph isomorphism detection: A decomposition approach. *IEEE Trans. Knowl. Data Eng.*, **12**(2), 307–323.
- Miyajima, K. and Ralescu, A. (1994). Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations. *Fuzzy Sets and Systems*, **2**(65), 225–236.
- Papadias, D. and Theodoridis, Y. (1997). Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science*, **11**(2), 111–138.
- Pham, T. V. and Smeulders, A. W. M. (2006). Learning spatial relations in object recognition. *Pattern Recognition Letters*, **27**(14), 1673–1684.
- Rendek, J., Masini, G., Dosch, P., and Tombre, K. (2004). The search for genericity in graphics

- recognition applications: Design issues of the qgar software system. In S. Marinai and A. Dengel, editors, *Proceedings of International Workshop on Document Analysis Systems*, volume 3163 of *Lecture Notes in Computer Science*, pages 366–377. Springer-Verlag.
- Reynolds, A. P., Richards, G., de la Iglesia, B., and Rayward-Smith, V. J. (2006). Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *Journal of Math. Model. Algorithms*, **5**(4), 475–504.
- Rohlf, F. J. and Fisher, D. R. (1968). Tests for hierarchical structure in random data sets. *Systematic Biology*, **17**(4), 407–412.
- Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**(1), 53–65.
- Rusiñol, M. and Lladós, J. (2006). Symbol spotting in technical drawings using vectorial signatures. In W. Liu and J. Lladós, editors, *Graphics Recognition. Ten Years Review and Future Perspectives*, volume 3926 of *Lecture Notes in Computer Science*, pages 35–46. Springer-Verlag.
- Rusiñol, M., Borràs, A., and Lladós, J. (2010). Relational indexing of vectorial primitives for symbol spotting in line-drawing images. *Pattern Recognition Letters*, **31**(3), 188–201.
- Saitta, S., Raphael, B., and Smith, I. F. (2007). A bounded index for cluster validity. In *Proceedings of International conference on Machine Learning and Data Mining in Pattern Recognition*, pages 174–187, Berlin, Heidelberg. Springer-Verlag.
- Salmon, J. P., Wendling, L., and Tabbone, S. (2007). Improving the recognition by integrating the combination of descriptors. *International Journal on Document Analysis and Recognition*, **9**(1), 3–12.
- Santosh, K. C., Lamiroy, B., and Ropers, J.-P. (2009). Inductive logic programming for symbol recognition. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 1330–1334.
- Santosh, K. C., Wendling, L., and Lamiroy, B. (2010). Unified pairwise spatial relations: An application to graphical symbol retrieval. In J.-M. Ogier, W. Liu, and J. Lladós, editors, *Graphics Recognition. Achievements, Challenges, and Evolution*, volume 6020 of *Lecture Notes in Computer Science*, pages 163–174. Springer-Verlag.
- Santosh, K. C., Lamiroy, B., and Wendling, L. (2011a). Dtw for matching radon features: A pattern recognition and retrieval method. volume 6915 of *Lecture Notes in Computer Science*, pages 249–260. Springer.
- Santosh, K. C., Lamiroy, B., and Wendling, L. (2011b). Symbol recognition using spatial relations. *Pattern Recognition Letters*, **33**(3), 331–341.
- Santosh, K. C., Lamiroy, B., and Wendling, L. (2013). Spatio-structural symbol description with statistical feature add-on. In Y.-B. Kwon and J.-M. Ogier, editors, *Graphics Recognition. New Trends and Challenges*, volume 7423 of *Lecture Notes in Computer Science*, pages 228–237. Springer.
- Sokal, R. R. and Rohlf, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *Taxon*, **11**(2), 33–40.
- Tabbone, S., Wendling, L., and Salmon, J.-P. (2006). A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding*, **102**(1), 42–51.
- Terrades, O. R., Valveny, E., and Tabbone, S. (2008). On the combination of ridgelets descriptors

- for symbol recognition. In W. Liu, J. Lladós, and J.-M. Ogier, editors, *Graphics Recognition. Recent Advances and New Opportunities*, volume 5046 of *Lecture Notes in Computer Science*, pages 40–50. Springer-Verlag.
- Terrades, O. R., Valveny, E., and Tabbone, S. (2009). Optimal classifier fusion in a non-bayesian probabilistic framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(9), 1630–1644.
- Tombre, K. (2010). Graphics recognition - what else? In J.-M. Ogier, W. Liu, and J. Lladós, editors, *Graphics Recognition. Achievements, Challenges, and Evolution*, volume 6020 of *Lecture Notes in Computer Science*, pages 272–277. Springer.
- Tombre, K. and Lamiroy, B. (2008). Pattern recognition methods for querying and browsing technical documentation. In *Proceedings of Iberoamerican Congress on Pattern Recognition*, pages 504–518. Springer-Verlag.
- Tombre, K., Ah-Soon, C., Dosch, P., Habed, A., and Masini, G. (1998). Stable, robust and off-the-shelf methods for graphics recognition. *Proceedings of International Conference on Pattern Recognition*, **1**, 406.
- Tooley, M. and Wyatt, D. (2008). *Aircraft electrical and electronic systems: principles, operation and maintenance*. Aircraft engineering principles and practice. Butterworth-Heinemann.
- Valveny, E. and Martí, E. (2003). A model for image generation and symbol recognition through the deformation of lineal shapes. *Pattern Recognition Letters*, **24**(15), 2857–2867.
- Wang, X. and Keller, J. (1999). Human-based spatial relationship generalization through neural/fuzzy approaches. *Fuzzy Sets and Systems*, **101**(1), 5–20.
- Wenyin, L., Zhang, W., and Yan, L. (2007). An interactive example-driven approach to graphics recognition in engineering drawings. *International Journal on Document Analysis and Recognition*, **9**(1), 13–29.
- Xiaogang, X., Zhengxing, S., Binbin, P., Xiangyu, J., and Wenyin, L. (2004). An online composite graphics recognition approach based on matching of spatial relation graphs. *International Journal on Document Analysis and Recognition*, **7**(1), 44–55.
- Yang, S. (2005). Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(2), 278–281.
- Yuen, P. C., Feng, G.-C., and Tang, Y. Y. (1998). Printed chinese character similarity measurement using ring projection and distance transform. *International Journal of Pattern Recognition and Artificial Intelligence*, **12**(2), 209–221.
- Zhang, D. and Lu, G. (2002). Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, **17**(10), 825–848.
- Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, **37**(1), 1–19.
- Zhang, W., Wenyin, L., and Zhang, K. (2006). Symbol recognition with kernel density matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(12), 2020–2024.