BoR: Bag-of-Relations for Symbol Retrieval

Santosh K.C., Laurent Wendling, Bart Lamiroy

HAL Id: hal-01025765

https://hal.inria.fr/hal-01025765

Submitted on 18 Jul 2014

BoR: Bag-of-Relations for Symbol Retrieval*

K.C. SANTOSH†

US National Library of Medicine (NLM), National Institutes of Health (NIH)

8600 Rockville, Bethesda, MD 20894, USA

LAURENT WENDLING

SIP – LIPADE, Université Paris Descartes (Paris V)

45, rue des Saints-Pères, 75270 Paris Cedex 06, FRANCE

BART LAMIROY

Université de Lorraine – LORIA (UMR-7503) Campus Scientifique

BP 239 – 54506 Vandoeuvre-les-Nancy Cedex, FRANCE

†Corresponding author

Abstract. In this paper, we address a new scheme for symbol retrieval based on bag-of-relations (BoRs) which are computed between extracted visual primitives (e.g. *circle* and *corner*). Our features consist of pairwise spatial relations from all possible combinations of individual visual primitives. The key characteristic of the overall process is to use topological relation information indexed in bags-of-relations and use this for recognition. As a consequence, directional relation matching takes place only with those candidates having similar topological configurations. A comprehensive study is made by using several different well known datasets such as GREC, FRESH and SESYD, and includes a comparison with state-of-the-art descriptors. Experiments provide interesting results on symbol spotting and other user-friendly symbol retrieval applications.

Keywords. Visual vocabulary · spatial reasoning · bag-of-relations · symbol localisation and retrieval.

1. Introduction

1.1 Motivation

Document analysis or processing is mainly related to both text and graphic separation, localisation and recognition (Fletcher & Kasturi, 1988). According to (Nagy, 2000), document analysis is related to document image analysis (DIA) since the much research has been concerned with document image interpretation such as segmentation, layout understanding and symbol recognition. Our work is to be positioned in this context, as we address the recognition of *graphical elements* and the localisation of significant or known visual parts of a document image. In other words, a graphical element can be defined as a symbol with a particular meaning in the context of a specific application domain (Lladós et al., 2002). Our goal is very similar to spotting, but we view this as a kind of retrieval (Delalandre et al., 2010; Qureshi et al., 2008; Tabbone et al., 2004) which is basically guided by user queries.

1.2 State-of-the-art

Graphics recognition has an extremely rich state-of-the-art literature in symbol recognition and localisation since the 70's (Bunke & Wang, 1997; Doermann & Tombre, 2014). In (Doermann & Tombre, 2014), a comprehensive overview of graphics recognition is provided. Existing approaches, specifically those based on feature based matching, can be split into three classes: statistical, structural and syntactic. The latter approach is not included, which is beyond the scope of the paper. Most methods are particularly suited for isolated line symbols, not for composed symbols connected to a complex environment (Cordella & Vento, 2000a;b; Lladós et al., 2001; Lladós et al., 2002).

1.2.1 Statistical approaches

In (Cordella & Vento, 2000a), shape analysis for symbol representation has been comprehensively addressed. It mainly refers to simple isolated $2D$ binary shapes. Under statistical approaches, global signal-based descriptors (Belongie et al., 2002; Kim & Kim, 2000; Tabbone et al., 2006; Yuen et al., 1998; Zhang & Lu, 2002; 2004) are usually quite fault tolerant to image distortions, since they tend to filter out small detail changes. This is unfortunately inappropriate in our context. Our context can be summarised as follows (FRESH dataset in Section 3): 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, they difficultly accommodate with connected or composite symbols. More precisely, the 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)). Despite those drawbacks, researchers have been integrating descriptors (Barrat & Tabbone, 2010; Salmon et al., 2007; Terrades et al., 2008) as well as 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 optimal set of features for each specific application.

1.2.2 Structural approaches

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 & Messmer, 1995; Conte et al., 2004), region adjacency graphs (RAG) (Lladós et al., 2001), constraint networks (Ah-Soon & Tombre, 2001) as well as deformable templates (Valveny & 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 (Bunke & Messmer, 1997; Foggia et al., 2014; Lladós & Sánchez, 2004; Robles-Kelly & Hancock, 2004). In this framework, another interesting example integrating shape descriptions with relations to form RAGs 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. Besides, it uses few model classes to localise symbols in technical documents. In the framework of stroke-based hand-drawn symbol recognition, prominent studies have been presented in (Kara & Stahovich, 2005; Lee et al., 2007). The first study is related to template-based matching and another one uses ARGs where the vertices represent geometric primitives like lines and arcs (based on their shapes) and the edges represent the geometric relationships between them. For matching, it is primarily based on graph matching or graph isomorphism detection presented in (Messmer & Bunke, 2000). The work is conceptually similar to (Xiaogang et al., 2004). These approaches perform well as long as vertices are well separated since they are taken from on-line strokes (*i.e.* vectorisation difficulties are avoided). Considering the time complexity issue, several works are more focused on computing symbol signatures by taking some regions of interest in the document image (Dosch & Lladós, 2004; Rusiñol & Lladós, 2006; Wenyin et al., 2007). These methods aim to provide faster matching in comparison to graph matching. But on the other hand, this methods are found to be limited since they make the strong assumption that the symbols always fall into a region of interest (ROI). Very recently, (Coustaty et al., 2011; Visani et al., 2011) introduced an interesting approach where a Galois lattice is used to classify structural signatures that are extracted using the Hough transform. These structural signatures are based on a topological graph, where there are only five topological relations computed between the segments (based on connected and disconnected topological configurations). As reported in the paper, the Galois lattice-based

classification its robust to noise. However, this may provide less consistent performance when symbols are found to be connected to other graphical elements or texts in the image.

1.3 Our contribution outline

Considering the problem of symbol localisation in real documents, composed of individual parts and constrained by spatial relations, global signal-based descriptors cannot be applied since they are, unfortunately, primarily designed for applications where line symbols are isolated. Thus, the problem is related to the segmentation/recognition paradigm (Yoon et al., 2001), where an accurate segmentation of meaningful parts is expected. In this context, one needs to be able to extract visual primitives like *saliency points*, *lines* and *arcs* and to formalise the possible links that exist between them to build a graph-like structure. Graph-based symbol recognition techniques are powerful but can suffer from time complexity issues. In our case, the use of *Bag-of-Relations* (BoR) indexing will reduce the execution time during the recognition process.

As shown in Fig. 1 and in previous works (Santosh, Lamiroy & Wendling, 2011; Santosh et al., 2014), our work[2] is based on extracted vocabulary that is semantically significant (i.e., visual primitives such as *circle*s and *corner*s). These visual primitives are then used for building BoRs based on their pairwise topological relations and directional relations. In other words, we integrate topological and directional relations for all possible combinations of the visual primitives that compose the symbol. Our bags correspond to topological relations between the visual primitives and directional relations are computed so that relation precision can further be exploited. For example, the sole topological relation when two primitives are *disconnected* does not convey any information about how they are oriented. This means that topological relations may not always be sufficient and therefore combining them with directional relations may prove useful for recognition. For indexing, the number of bags is limited to the number of possible topological relations, regardless of shape, size and number of the visual primitives that compose the symbol, and directional relations are computed and stored. Consequently, for recognition, directional relation matching takes place only with those which share similar topological and vocabulary type information. This not only simply reduces the computational complexity in matching but also avoids irrelevant relation matching, thanks to vocabulary categorisation.

The paper addresses some of the shortcomings of previous work (Santosh, Lamiroy & Wendling, 2011) which uses spatial relations and extends its scope in such a way that it can be used in a context of flexible querying, retrieval and localisation of symbols. In this previous work, visual elements are grouped by type. While thos avoids the NP-hardness of the underlying graph matching problem, it requires at least two different types of visual primitives in a symbol to compute the needed spatial relations. It is therefore inappropriate for symbols having only a single vocabulary type (e.g. four *corners* from a rectangle-shaped symbol) regardless how many visual primitives it contains. Furthermore, in (Santosh et al., 2014) we have improved the overall recognition performance extending the used vocabulary

---

[2] This manuscript is the extension of previously presented work presented in (Santosh et al., 2013).
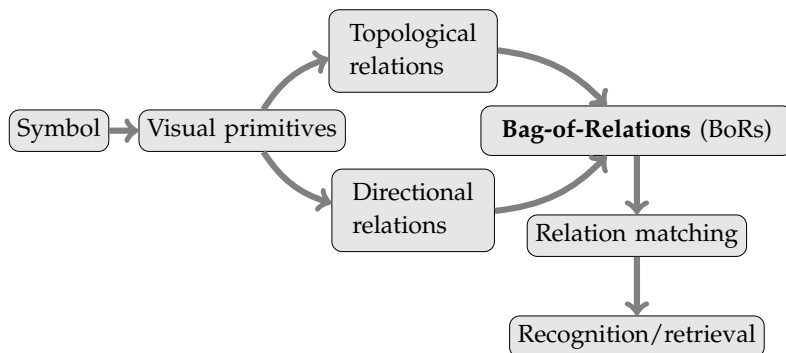
Fig. 1. An architecture of the proposed method. It uses spatial relations between the visual primitives that compose a symbol, as the candidates of BoRs.

primitives via generic unsupervised clustering. The advancement presented in this paper consists in computing all possible relations that exist between individual visual primitives, rather than relying on classes of primitives. However, since computing all possible relations that exist between individual visual primitives is computationally expensive we reduce the execution by using *Bag-of-Relations* (BoR) indexing. This is the core contribution of this paper.

1.4 Organisation of the paper

The rest of the paper is organised as follows. In Section 2, we explain our method by describing BoRs and recognition techniques. Full experimental results are reported and analysed in Section 3. In Section 4, a user-friendly symbol localisation/retrieval application is provided. The paper is concluded in Section 5.

2. Bag-of-relations

In this section we explain how visual primitives are extracted from the symbol in Section 2.1. In Section 2.2, we show how this can be used for categorisation. As soon as we have categorised visual primitives, we discuss the details of how directional relations fit in, in Section 2.3. After that, we explain how relations are used for recognition and retrieval in Section 2.4. In Section 2.5, we discuss time complexity.

2.1 Visual vocabulary

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; Lamiroy & Guebbas, 2010; Rendek et al., 2004). Our vocabulary set consists of *thick* (filled) components, *circles*, *corners* and *extremities*. Fig. 2.1 shows some examples.
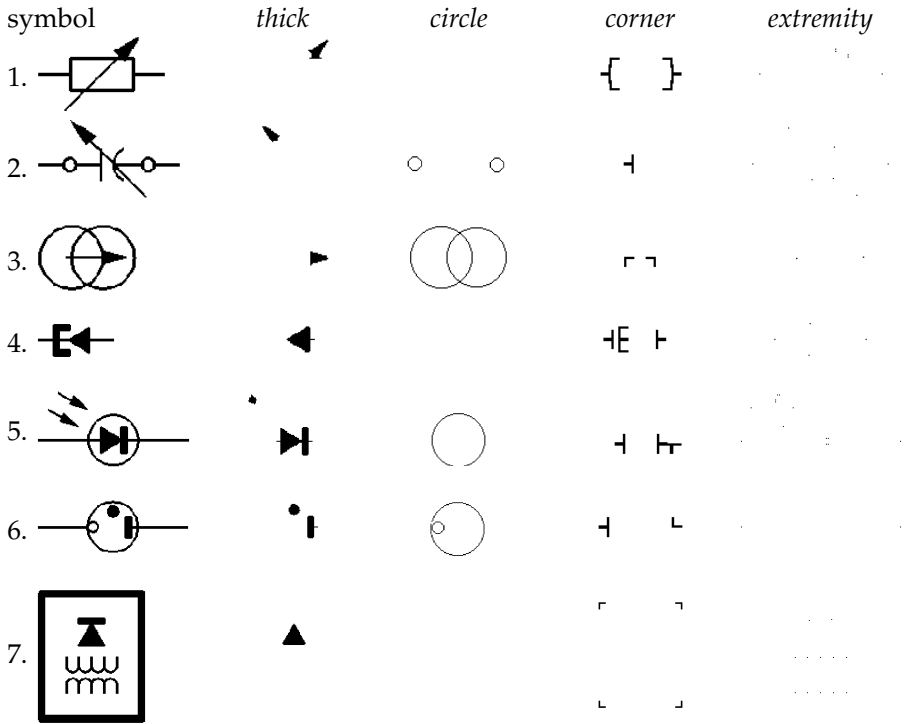
Fig. 2. Visual vocabulary from a few symbols.

1. *thick* primitive

   We employ straightforward thin/thick stroke 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.

   Our current implementation is based on (Dosch et al., 2000; Rendek et al., 2004) and uses a straightforward histogram high-pass filter on the line thickness in the document image.

2. *circle* primitive

   We use the algorithm as described in (Lamiroy & Guebbas, 2010) based on Ransac minimisation.

3. *corner* primitive

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

4. *extremity* primitive

   We detect loose ends from the image skeleton pixels, where there is only a unique

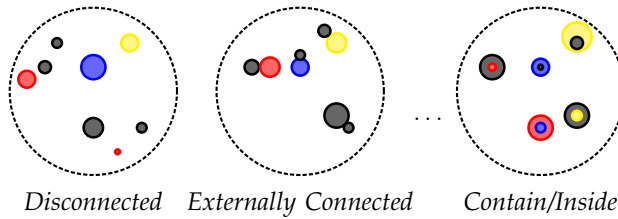*Disconnected*   *Externally Connected*   *Contain/Inside*

Fig. 3. Our bag-of-relations (BoRs) model. Each item in every bag represents a visual primitive and its colour represents its vocabulary type.

neighbouring pixel connecting to the main skeleton, which itself is connected by a unique neighbouring pixel.

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

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

It is to be noted that the choice of vocabulary types has an impact on time complexity and recognition rate. Section 3.1.3 shows, for instance, how *extremities* influence the results.

2.2 Vocabulary categorisation

In this section, we first give an overview of how we represent vocabulary categorisation via topological relations. We then provide some real-world examples to illustrate the overall idea.

2.2.1 Representing vocabulary categorisation

Any symbol $\mathcal{S}$ is decomposed into a variable number p of visual primitives, each of which belongs to a vocabulary type $\mathbb{T}_t$ (in our case $1 \leq t \leq 4$). For any vocabulary type $\mathbb{T}_t$, there are m_t visual primitives,

$$\mathbb{T}_t = \{\wp_i^t\}, i = [1, \ldots, m_t] \text{ and } p = \sum_t m_t. \tag{2}$$

Any pair of primitives $(\wp_1, \wp_2)$, as illustrated in Fig. 3, can be represented by both the vocabulary types each part belongs to (represented by their color) and by the topological relation that characterises them:

1. *disconnected* (DC),
2. *externally connected* (EC),
3. *overlap* (O),
4. *contain/inside* (Cn/I),
5. *cover/covered by* (Cr/CB), and
6. *equal* (EQ).

This means that in general, visual primitives are categorised into six topological relations. For simplicity, we rewrite such a set of topological relations as

$$\text{voc. categorisation} = \{\mathcal{C}_{\Bbbk}\}, \Bbbk = [1, \ldots, 6], \tag{3}$$

preserving the label ordering $\{\mathcal{C}_{DC}, \mathcal{C}_{EC}, \mathcal{C}_O, \ldots, \mathcal{C}_{EQ}\}$.

2.2.2 Computing topological relations

To obtain the topological relation between two primitives $\mathcal{T}(\wp_1, \wp_2)$, we use the 9-intersection model (Egenhofer & Franzosa, 1991; Egenhofer & Herring, 1991; Güting, 1988) relative to the boundaries ($\partial *$), interiors ($*^o$) and exteriors ($*^-$) of $\wp_1$ and $\wp_2$:

$$\mathcal{T}(\wp_1, \wp_2) = \begin{bmatrix} \wp_1^o \cap \wp_2^o & \wp_1^o \cap \partial\wp_2 & \wp_1^o \cap \wp_2^- \\ \partial\wp_1 \cap \wp_2^o & \partial\wp_1 \cap \partial\wp_2 & \partial\wp_1 \cap \wp_2^- \\ \wp_1^- \cap \wp_2^o & \wp_1^- \cap \partial\wp_2 & \wp_1^- \cap \wp_2^- \end{bmatrix}. \tag{4}$$

Their definitions use basic set operations like $=, \neq, \subseteq$ and $\cap$ (Güting, 1988). For example,

- *equal*$(\wp_1, \wp_2) := \text{points}(\wp_1) = \text{points}(\wp_2)$;
- *disconnected*$(\wp_1, \wp_2) := \text{points}(\wp_1) \neq \text{points}(\wp_2)$ or $\text{points}(\wp_1) \cap \text{points}(\wp_2) = \varnothing$;
- *inside*$(\wp_1, \wp_2) := \text{points}(\wp_1) \subseteq \text{points}(\wp_2)$; and
- *intersects*$(\wp_1, \wp_2) := \text{points}(\wp_1) \cap \text{points}(\wp_2) \neq \varnothing$.

Since the *intersects* definition covers both *equal* and *inside*, they must be separated. Therefore, the previous definitions have been augmented with the consideration of boundary and interior so that the *overlap* and *externally connected* can be distinguished (Pullar & Egenhofer, 1988):

- *overlap*$(\wp_1, \wp_2) := \partial\wp_1 \cap \partial\wp_2 \neq \varnothing \ \& \ \wp_1^o \cap \wp_2^o \neq \varnothing$; and
- *externally connected*$(\wp_1, \wp_2) := \partial\wp_1 \cap \partial\wp_2 \neq \varnothing \ \& \ \wp_1^o \cap \wp_2^o = \varnothing$.

Therefore, the topological relation $\mathcal{T}(\wp_1, \wp_2)$ provides a boolean value for each of the elements of the matrix shown in (4). It is straightforward to combine these elements to obtain $\{\mathcal{C}_{\Bbbk}\}$.

Based on the symbols in Fig. 2, the following samples illustrate how the pairwise topological relations are expressed.

1.   from symbol 1:

In this case, the symbol consists of a single *thick* ✈ and several *corner* visual primitives. The *thick* primitive is *externally connected* to a *North-East corner* ⌐ i.e, EC(✈,⌐). The remaining *corner*s are *disconnected* to the *thick*. In a similar manner, there exist *externally connected* and *disconnected* intra-*corner* topological configurations. Finally, we have two different topological categories (and thus bags): *disconnected* and *externally connected*.
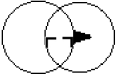
2.   ○ ⊣   ○   from symbol 2:

In this case, all possible combinations of visual primitives are found to be in *disconnected* configurations except two neighbouring *corners*: *South-East* ⌟ and *North-East* ⌐ . As a consequence, we have two different bags: *disconnected* and *externally connected*.

3.   (⊢→)   from symbol 3:

In this example, *circles* ( ) are *overlapped*. When *thick* ▶ is taken into account, two different topological relations i.e., *externally connected* (with *circle* on the left) and *contain/inside* (with *circle* on the right) are found. Similarly, *North-East* corner ⌐ is *externally connected* with *thick*, *North-West* corner ⌐ is *disconnected* with *thick*, and corners are *disconnected*. Both corners are *covered by* circles. On the whole, we have four bags: *overlap*, *externally connected*, *cover/covered by* and *contain/inside*.

In our study, we note that all predefined categories may not be used since all topological relations may not be present between the visual primitives belonging to a single symbol.

As soon as we have categorised visual primitives via topological relations, we compute the corresponding directional spatial relation metrics between all pairs.

2.3 Spatial reasoning

Like in Section 2.2, we first provide spatial reasoning representation via directional relations, and then examples to illustrate the idea.

2.3.1 Representing spatial reasoning

Until now, we have categorised visual primitives by using their topological relations in addition to their vocabulary type information i.e., $\left(\wp_i^t, \wp_j^{t'}\right) \mapsto \left(\left(\mathbb{T}_t, \mathbb{T}_{t'}\right), \mathcal{T}\left(\wp_i, \wp_j\right)\right)$, where $\mathcal{T}(,)$ refers to visual primitives' categorisation $\Bbbk$. Based on this, we then compute directional relations $\mathcal{D}$, which is represented by, $\left\{\mathcal{D}_{\Bbbk}\left(\wp_i^t, \wp_j^{t'}\right)\right\} \forall i, j$ in any studied symbol.

2.3.2 Computing directional relations

Pairwise spatial relations are often expressed by taking one of the objects as a reference. For example, $\wp_1$ is to the *right* of $\wp_2$: $right(\wp_1, \wp_2)$, where $\wp_2$ is referenced. In our context, since the number of visual primitives is not always the same, it is difficult to fix a particular primitive as a reference. To avoid such a difficulty, we first set up a unique reference $\mathbb{R}$ from each pair. $\mathbb{R}$ is either the common portion of two neighbouring sides in the case of *disconnected* primitives or the intersection in the case of *overlapping*, *equal* or otherwise *connected* primitives that are maintained with minimum bounding boxes. Depending on the topological relations, $\mathbb{R}$ can be

- $2D$: rectangle or area that are computed from all topological relations except *externally connected*;
- $1D$: line that can be vertical and horizontal, which are found to be happened in *disconnected* and *externally connected* relations; and
- $0D$: a single point that is found only in *externally connected* relation.

Then, we compute directional relations with respect to the unique reference, thus avoiding potential ambiguity. To compute directional relations between the primitives in all categories, multiple options are available. In this paper, we will primarily use

1. projection-based minimum boundary rectangle (MBR) model (Papadias & Theodoridis, 1997) and
2. radial line model (Santosh, Lamiroy & Wendling, 2011).

Since our approach can use any spatial relation model, we will report the comparative performance of some common state-of-the-art spatial relations models in our experiments in Section 3.

Minimum boundary rectangle (MBR) model

Given a unique reference point set $\mathbb{R}$, cardinal relations, based on projection are designed for developing 9-directional relation predicates:

$$\begin{bmatrix} \text{left top (LT)} & \text{top (T)} & \text{right top (RT)} \\ \text{left (L)} & \text{middle (M)} & \text{right (R)} \\ \text{left bottom (LB)} & \text{bottom (B)} & \text{right bottom (RB)} \end{bmatrix}.$$

Computing the relational matrix $\mathcal{M}$ simply consists in checking whether the part of the studied visual primitive $\wp_X$ is found to be present in one or more of the directional spaces made by $\mathbb{R}$ via orthogonal projection. In our case, various cases can be studied since we have several $\mathbb{R}$ variants. Unlike $2D$ projection, in $1D$ (line) projection, a few directional spaces are omitted. A single point i.e., $0D$ projection provides the standard four directional spaces. In all
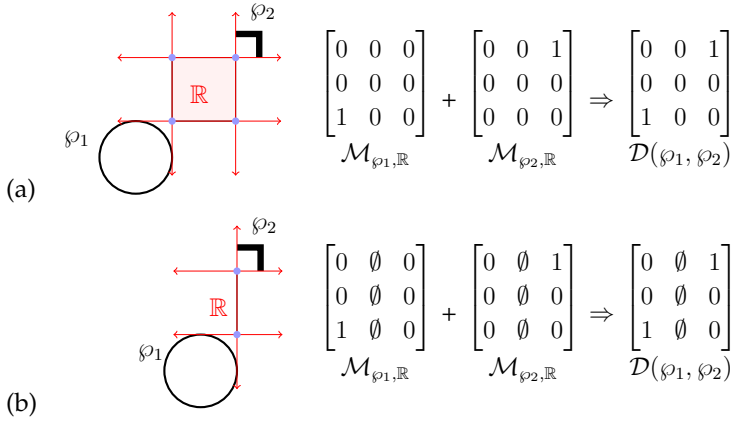
Fig. 4. A *disconnected* pair of $\wp_1^{circle}$ and $\wp_2^{corner}$ including unique reference point set $\mathbb{R}$ and directional relations using MBR with respect to $\mathbb{R}$. MBR is applied for both visual primitives and combined together.

cases, we keep the standard 3×3 matrix by using '$\varnothing$' for the unused spaces. Considering two visual primitives $\wp_1$ and $\wp_2$, we consider the directional relation $\mathcal{D}$ as

$$\mathcal{D}(\wp_1, \wp_2) = \{\mathcal{M}_{\wp_1, \mathbb{R}}, \mathcal{M}_{\wp_2, \mathbb{R}}\}. \tag{5}$$

Instead of using them separately, we combine them together to form a single relation matrix, as shown in Fig. 4 for a *disconnected* topological relation.

Radial line model (RLM)

For a given reference point $\mathbb{R}_{p_c}$ taken to be the centroid of $\mathbb{R}$, we cover the surrounding space by rotating a radial line $line(\mathbb{R}_{p_c}, \theta)$ over a cycle. While rotating, we consider the sector $sector(\mathbb{R}_{p_c}, \theta_{j,j+1})$ delimited by two consecutive radial lines at θ_j and θ_{j+1}. Therefore, we are having a regular radial intervals of $\theta_{j,j+1} = \frac{2\pi}{s}$, where s is the number of sectors. In each sector, we compute a histogram $\mathcal{H}$, counting the percentage of pixels of the studied visual primitives lying in it,

$$\mathcal{H}_{\wp_1, \mathbb{R}_{p_c}} = \left[\frac{sector(\mathbb{R}_{p_c}, \theta_{j,j+1}) \cap \text{points}(\wp_1)}{\text{points}(\wp_1)} \right]_{j=0,\ldots,s-1}. \tag{6}$$

Fig. 5 shows an illustration of this process. It shows how a sector can be built using two consecutive radial lines and how the histogram can be computed in that sector. At this point, it is important to notice that the value of $\theta_{j,j+1}$ determines the precision (or accuracy) of the histogram. The lower $\theta_{j,j+1}$, the better the performance (Santosh, Lamiroy & Wendling, 2011). Considering both visual primitives $\wp_1$ and $\wp_2$, directional relation $\mathcal{D}$ is computed by combining their corresponding histograms,

$$\mathcal{D}(\wp_1, \wp_2) = \{\mathcal{H}_{\wp_1, \mathbb{R}_{p_c}}, \mathcal{H}_{\wp_2, \mathbb{R}_{p_c}}\}. \tag{7}$$
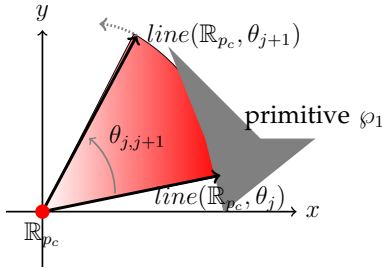
(a) Sector made by two consecutive radial lines in $\mathbb{R}^2$ space.

(b) A single histogram value in a defined sector.

Fig. 5. An illustration of how (a) sector is made using two consecutive radial lines at θ_j and θ_{j+1}, and (b) histogram computation in that sector.

As before (in the MBR model), both histograms are combined to make a single vector. Fig. 6 shows the results on the examples from Fig. 4.

Instead of just looking at how relation models work, Figs. 4 and 6 show how discriminant they are, even when the visual primitive pairs have similar topological relations. In contrast to the MBR model, RLM provides a more detailed relational signature.

2.4 Recognition

We have defined each visual primitive's type, as well as their pairwise topology and directional relation signatures. This means that each symbol can be represented by BoRs where directional relations are stored in the bags that are derived from their topological relations.

Based on this symbol representation, we can describe our recognition method. Symbol recognition is based on the matching between corresponding pairs of visual primitives. In order to reduce combinatorial complexity we only consider configurations where matching candidate pairs share identical values for their vocabulary types and topological relations, thanks to the indexing of topological bags.

In what follows, we first provide an idea of how relation matching happens to compute the distance between two symbols, followed by the whole symbol recognition/retrieval process.

2.4.1 Relation matching

Consider a symbol $\mathscr{S}$ having a set of visual primitives $\{\wp_i^t\}$ (meaning primitive i belongs to vocabulary type $\mathbb{T}_t$). All possible pairs of primitives can be categorised in one of the topological configuration categories $\{\mathcal{C}_\Bbbk\}$ and can be attributed with directional relation signatures $\left\{\mathcal{D}_\Bbbk(\wp_i^t, \wp_{i'}^{t'})\right\}$.
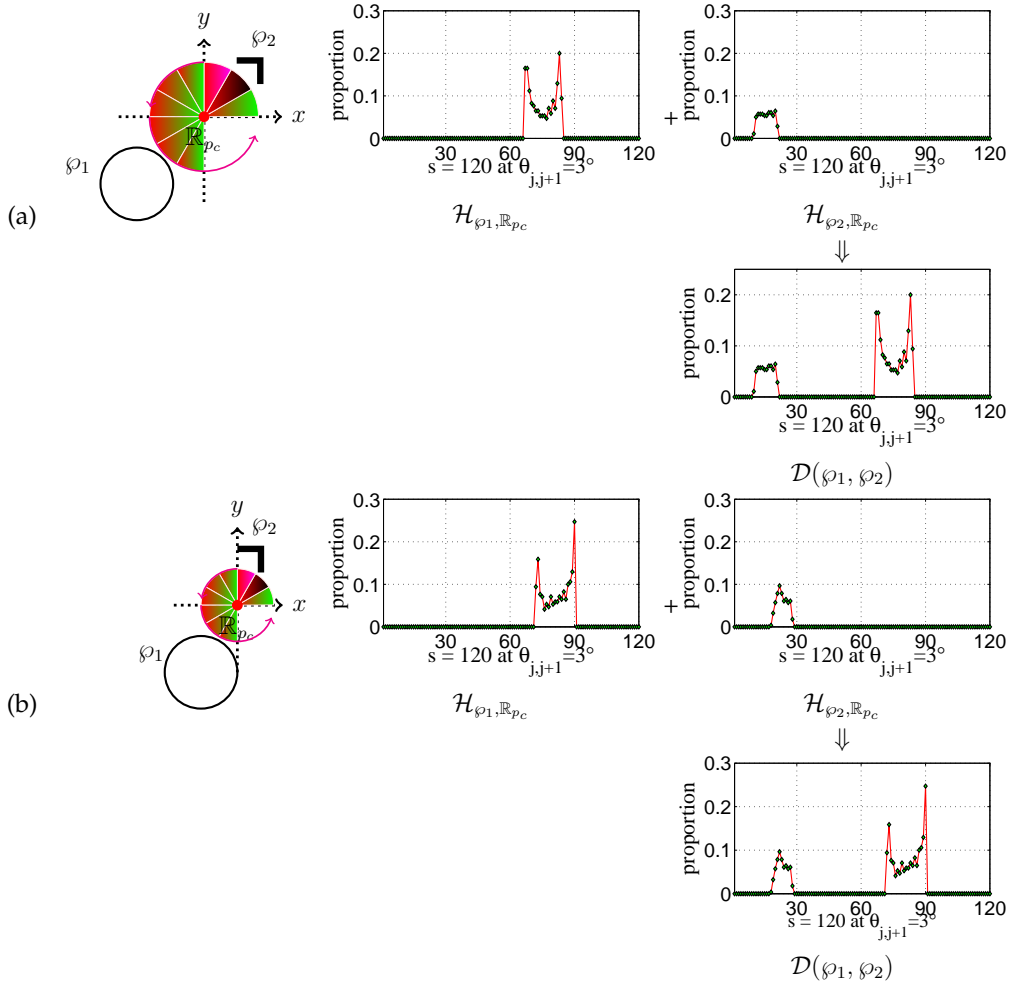
Fig. 6. Two *disconnected* pairs of visual primitives $\wp_1^{circle}$ and $\wp_2^{corner}$ and directional relational histograms using RLM with respect to $\mathbb{R}_{p_c}$. RLM is applied for both visual primitives and the obtained histograms are merged.

Given a query symbol $\mathscr{S}^q$ and a database symbol $\mathscr{S}^d$, the distance between them can be computed as follows.

For each query pair $(\wp_i^t, \wp_{i'}^{t'})$ of $\mathscr{S}^q$, we select the bag of database pairs $(\wp_{\hat{i}}^t, \wp_{\hat{i}'}^{t'})$ from $\mathscr{S}^d$ belonging to exactly the same topological relation category $\mathcal{C}_\Bbbk$. Within this category, we match query and database pairs by computing the distance between their directional relation signatures $\mathcal{D}(,)$. Distances are computed only between pairs sharing identical vocabulary types. The distance between $(\wp_i^t, \wp_{i'}^{t'})$ and $(\wp_{\hat{i}}^t, \wp_{\hat{i}'}^{t'})$ is defined as $\delta\left(\mathcal{D}_\Bbbk^q(\wp_i^t, \wp_{i'}^{t'}), \mathcal{D}_\Bbbk^d(\wp_{\hat{i}}^t, \wp_{\hat{i}'}^{t'})\right)$, where $\delta(a, b) = ||a - b||$. To illustrate this, Fig. 7 shows an example. In this example, a query
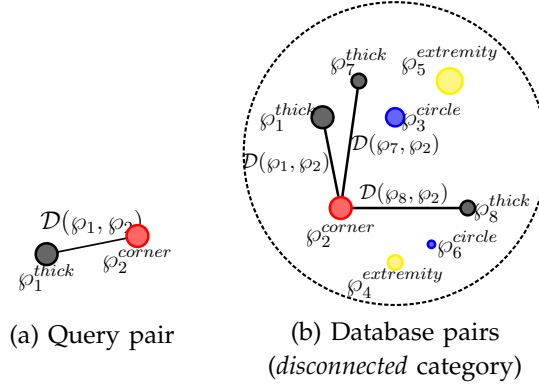
Fig. 7. Relation matching example in a *disconnected* category. Each item represents a visual primitive and colour represents its vocabulary type. For relation matching, visual primitives are only chosen if they share exactly similar vocabulary type information with the query pair.

pair (i.e., *thick* and *corner*) is *disconnected*. Therefore, the bag sharing identical topological relations is chosen from the database, where three *thick* primitives $\wp_1$, $\wp_7$ and $\wp_8$, and a single *corner* primitive $\wp_2$ are selected. Altogether four primitives are paired in accordance with the query pair. To find the best candidate, we then take the minimum distance of the three possible pairs if database primitives are extracted from the same symbol. If not, we keep them separately.

The closest matching distance pairs will be $\arg\min_{\hat{i}\hat{i}'} \delta(,)$ and for all query pairs $pair_{ii'}$ in a particular category $\Bbbk$, distance can be aggregated as,

$$\Delta_{\Bbbk}(\mathscr{S}^q, \mathscr{S}^d) = \sum_{ii'} \min_{\hat{i}\hat{i}'} \delta\left(\mathcal{D}^q_{\Bbbk}(\wp^t_i, \wp^{t'}_{i'}), \mathcal{D}^d_{\Bbbk}(\wp^t_{\hat{i}}, \wp^{t'}_{\hat{i}'})\right). \tag{8}$$

Finally, taking into account all pre-defined bags/categories, the total distance between two symbols can be expressed as,

$$\text{Dist.}(\mathscr{S}^q, \mathscr{S}^d) = \sum_{\Bbbk} \Delta_{\Bbbk}(\mathscr{S}^q, \mathscr{S}^d). \tag{9}$$

2.4.2 Symbol recognition and ranking

The Dist.$(,)$ of course, conveys how similar/dissimilar a database symbol is with respect to a query. In order to rank the similarity from 0 to 1, for any query q, we normalise Dist.$(\mathscr{S}^q, \mathscr{S}^d), d \in \{1, \ldots, \mathbb{D}\}$ with $\overline{\text{Dist.}}(,) = \frac{\text{Dist.}(,) - \text{Dist.}^{min.}(,)}{\text{Dist.}^{max.}(,) - \text{Dist.}^{min.}(,)}$. We then express similarity as,

$$\text{Similarity}(\mathscr{S}^q, \mathscr{S}^d) = 1 - \overline{\text{Dist.}}(\mathscr{S}^q, \mathscr{S}^d) \simeq \begin{cases} 1 & \text{the closest candidate and} \\ 0 & \text{the most dissimilar candidate.} \end{cases} \tag{10}$$

Database symbols are therefore, ranked based on the decreasing order of similarity. In our experiments, we will distinguish "recognition" (search for the closest candidate) from "retrieval" (where closest candidates are retrieved for a given short-list).

2.5 Time complexity

For a symbol $\mathscr{S}$, there are $p = \sum_t m_t$ (*cf.* Eq. (2)) number of primitives, where m is the number of primitives in a specific vocabulary type $\mathbb{T}_t$ defined in Eq. (1). To handle all possible pairs of primitives, one requires to compute r number of relations i.e., $r = \frac{p(p-1)}{2}$. In more detail, considering vocabulary categorisation (i.e., six different bags), we have then $r = \sum_{\Bbbk} r_{\Bbbk}$, where $r_{\Bbbk}$ is the number of relations from $p_{\Bbbk}$ number of primitives in any particular category $\Bbbk$.

Based on this, in what follows, we address the computational complexity for both learning and recognition modules.

1. Learning.
   To learn a symbol, since we use all possible relations that exist between visual primitives, time complexity $O\left(\frac{p(p-1)}{2}\right) < O(p^2)$, which is equivalent to the number of relations. This holds for both topological and directional relations, which is computationally expensive.

2. Recognition.
   For recognition, relation matching does not follow the similar computational complexity as in learning module. Since we use vocabulary type information in all categories (bags), we can truncate the number of relation matchings for every query pair. Consider there are $\hat{p}_{\Bbbk}$ number of primitives – sharing exactly similar vocabulary type information as in query pair – in any particular category $\Bbbk$, time complexity will then be $\sum_{\Bbbk=1}^{6} O\left(\frac{\hat{p}_{\Bbbk}(\hat{p}_{\Bbbk}-1)}{2}\right)$, and $\hat{p}_{\Bbbk} << p_{\Bbbk}$. Fig. 7 illustrates it with an example where only three *disconnected* pairs (from three visual primitives) are used to match with a query based on their vocabulary type information. Therefore, considering a query symbol $\mathscr{S}^q$ and a database symbol $\mathscr{S}^d$, total time complexity can be expressed as, $\sum_{\Bbbk} O\left(\frac{p_{\Bbbk}^q(p_{\Bbbk}^q-1)}{2} \times \frac{\hat{p}_{\Bbbk}^d(\hat{p}_{\Bbbk}^d-1)}{2}\right)$. In our case, since not all bags are used for all symbols (*cf.* Section 2.2), execution time can further be reduced. However, the asymptotic complexity over all known symbols remains $\sum_{D} \sum_{\Bbbk} O\left(p^4\right)$.

3. Experiments

3.1 Experimental setup

In this section, datasets and the way we use ground-truth will be discussed first, followed by the evaluation metric used for measuring the performance of the methods. We also perform a simple experiment to see whether we can optimise the performance based on the selection of vocabulary types. At the end of the section, some relevant works are discussed based on the previous studies.
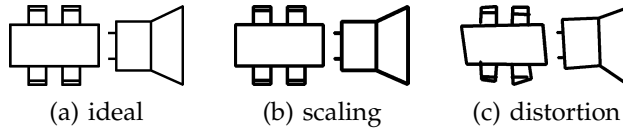
Fig. 8. Examples of graphical symbols from GREC dataset (GREC, 2003) from three different sets. In each set, two samples are shown.

3.1.1 Datasets

To validate the proposed BoRs method, we have used the following datasets:

1. GREC dataset (GREC, 2003),
2. FRESH dataset (Tooley & Wyatt, 2008), and
3. SESYD datasets[9] (Delalandre et al., 2010).

GREC 2003. In GREC dataset, there are altogether 50 different models. These model symbols are grouped into three sets, containing 5, 20 and 50 model symbols. In our test, ideal, scaling and distortion subsets are taken. Each model symbol has 5 test images in each subset besides for the "ideal" subset. Ideal test images are directly taken from the set of model symbols and therefore the test is just aimed to evaluate scalability. Since vectorial distortion works only on symbols with straight lines, and not with arcs, it is limited to 15 model symbols, and from which two sets are made respectively with 5 and 15 model symbols. Fig. 8 shows a few samples of it.

FRESH. FRESH dataset is composed of roughly 500 different known symbols, some of which come from (Santosh, Lamiroy & Wendling, 2011; Santosh et al., 2014; Tooley & Wyatt, 2008). The 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. Our dataset is completely unlabelled and imbalanced i.e., neither ground-truth is given nor identical number of similar symbols exist for all queries. For ground-truth formation, we have proceeded by using 6 human volunteers for validation, but by taking care of eliminating subjective bias. Human evaluators have chosen the candidates which have similar visual overall appearance or which have 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 human evaluator has selected the same result among other similar items. Fig. 9 gives an overview of the graphical symbols. For instance, for *query 1*, evaluators have provided a list of symbols which they consider visually close, or containing parts that are visually close. The evaluators were not required to provide any ranking order nor degree of visual resemblance.

SESYD. From this collection, we have taken two different datasets:

---

[9] http://mathieu.delalandre.free.fr/projects/sesyd/
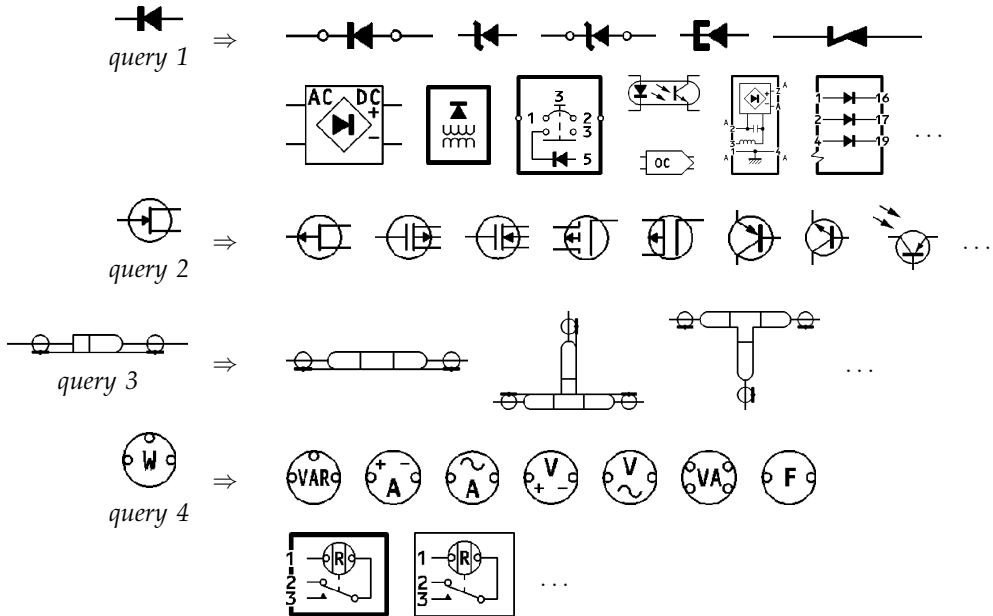
Fig. 9. Some examples of electrical symbols (Tooley & Wyatt, 2008). For every query symbol: *query 1* to *query 4*, a few relevant symbols are enlisted based on human evaluation. It consists of both linear as well as symbols in the composite form.

1. bags of symbols (BoS), and
2. electrical diagrams (ED).

From the BoS dataset, we have used 50 different models. These symbols are grouped into two sets, containing 25 and 50 model symbols. In our test, ideal and scaled models are taken. In both ideal and scaling subsets, there are 100 test images. A single database image contains around 10 symbols, which are randomly positioned without any connection, and are scaled by using different parameters. Fig. 10 shows a few examples.

From the ED dataset, we have used 10 models. In our test, three different subsets are taken from the pool, where each subset is composed of 100 images. The maximum number of symbols in a single database image is 26 and the minimum is 7 (14 symbols per image in average). Fig. 11 shows a few examples of it.

3.1.2 Evaluation metric

In case of datasets that are fully balanced and are composed of labelled ground-truths, it is rather straightforward to use recognition rate.

In case of datasets where the number of similar symbols (i.e., ground-truth) varies a lot from one query to another (imbalanced but labelled ground-truths), we use retrieval

Fig. 10. Examples illustrating three different test images from the 'bags of symbols' dataset (Delalandre et al., 2010).

efficiency (Kankanhalli et al., 1995) as a measure for retrieval quality. For a chosen query and for a fixed number of K returned symbols, it can be expressed as,
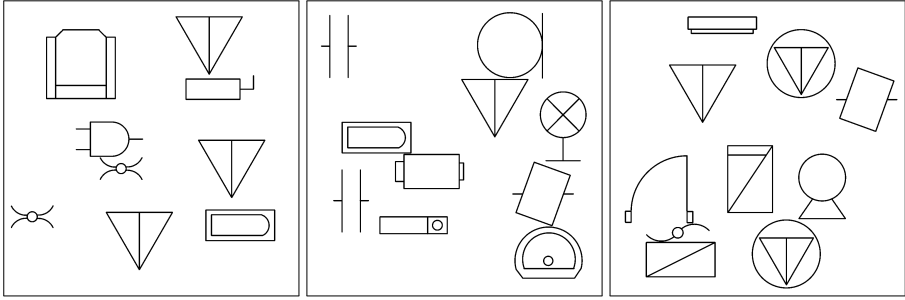
$$\eta_K = \begin{cases} n/N \text{ if } N \leq K \\ n/K \text{ otherwise,} \end{cases} \tag{11}$$

where n is the number of returned relevant symbols and N the total number of relevant symbols in the dataset. Note that η_K computes the traditional recall if $N \leq K$ and computes precision otherwise. The average retrieval efficiency curve is not biased even though the number of ground-truths varies from one query to another, while it happens when computing precision in case $N < K$. Therefore, retrieval efficiency is found to be the combined version of traditional precision and recall measure.

3.1.3 Vocabulary selection

Using visual primitives that are semantically meaningful can make a big performance difference. In this study, while considering the influence of time complexity, we aim to see whether *extremities* play an important role in recognition. The motivation behind this test is due to the extraction of large number of *extremities* even from a single image, in comparison to other visual primitives. Further, unlike other visual primitives, *extremity* alone does not provide any clue about visual recognition.

To handle this, we perform two different experiments:

1. 'with *extremities*' (WE) and

2. 'without *extremities*' (WoE).

In this test, we employ two different spatial relation models (MBR and RLM as discussed in Section 2.3) to avoid bias that can be possible from a single test. Table 1 shows the average recognition rates for both WE and WoE tests by using GREC 2003 dataset. In Table 1, we note that the WE provides better results (i.e., recognition rate of 97% from RLM) but, maintains marginal difference with WoE test. Therefore, the remaining test will include *extremities*

(a) Model symbols
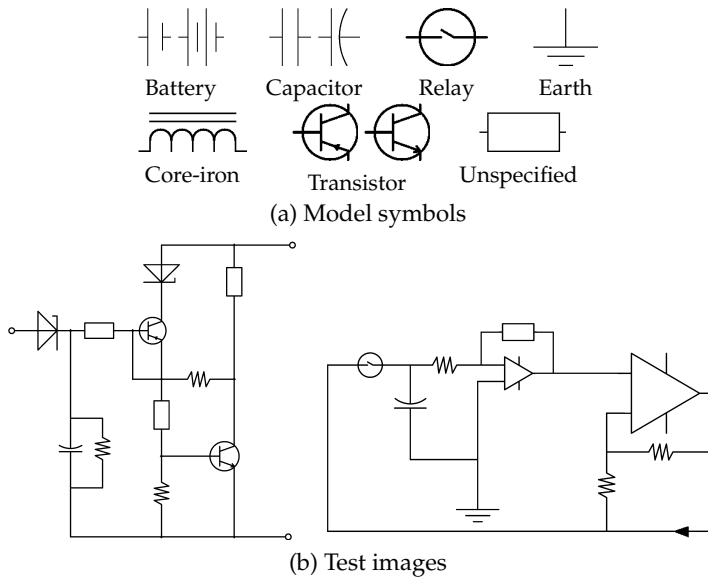
(b) Test images

Fig. 11. Examples illustrating (a) 10 model (query) symbols and (b) two test images from 'electrical diagram' dataset (Delalandre et al., 2010).

Table 1. Average recognition rate ('with *extremities*' (WE) and 'without *extremities*' (WoE)) in % using GREC 2003 dataset.

	ideal			scale		distort		Average
Bag-of-Relations	set 1	set 2	set 3	set 1	set 2	set 1	set 2	
MBR (WE)	100	95	96	85	84	85	82	90
MBR (WoE)	100	85	92	85	78	85	78	86
RLM (WE)	100	100	100	100	92	100	88	97
RLM (WoE)	100	95	97	100	89	100	84	95

even though it increases time complexity since our idea is to achieve the maximum possible recognition/retrieval performance. Fig. 12 shows a few samples where the use of *extremities* advances the overall recognition performance. But, considering the symbols in the composite form where symbols are connected with other graphical elements, *extremities* that are extracted from connecting terminals will be missed. In this context, *extremities* are still appearing in the symbols themselves (such as $\overline{\overline{\text{mm}}}$, $\perp$ and $\otimes$).

3.1.4 Related work

In our comparison, we will study the following state-of-the-art methods. They are grouped into three different sets:

1. spatial relation models;
2. global signal-based descriptors; and

Fig. 12. Examples (from GREC 2003 dataset) where *extremities* can advance recognition.

3. pixel-based approaches.

Other than RLM, we have used very common spatial relation models from the state-of-the-art: cone-shaped (Miyajima & Ralescu, 1994), angle histogram (Wang & Keller, 1999) and MBR (Papadias & Theodoridis, 1997). Our aim is to show that our proposed method is independent of the actual relation model used.

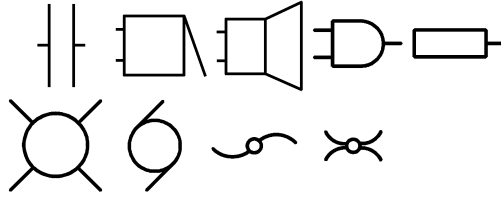We have selected a set of major global signal-based shape descriptors, ZM (Kim & Kim, 2000), GFD (Zhang & Lu, 2002), SC (Belongie et al., 2002), and $\mathcal{R}-$signature (Tabbone et al., 2006). For GFD, we have tuned the parameters, and selected those values for radial and angular frequency that achieved the best recognition performance on our dataset: radial frequency 6 and angular 15. For SC, only 70 (and maximum 100) sample points have been selected because of the presence of smaller size images in our database. In case of ZM, we have used 36 *zernike* functions of order less than or equal to 7. Also, we have taken radon image intensity over the projecting angle $[0, \pi[$ by default for $\mathcal{R}-$signature.

We have also employed two 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). In SIHA, two different length-ratio and angle-ratio histograms are taken from every two pixels in reference to a third pixel from the skeleton image. In KDM, skeleton symbols represent as $2D$ kernel densities and their similarity is measured by the *Kullback-Leibler divergence*.

Our aim is not to re-run all state-of-the-art methods on all test datasets, but to take those validations and thus conclusions made in the previous work (Santosh, Lamiroy & Wendling, 2011). In case of spatial relation models, the best performers are RLM and MBR. From the global signal-based descriptors, GFD seems to be performing the best among all tested global signal-based descriptors. But, the differences are marginal when we consider isolated symbol recognition. In case of pixel-based approaches, we observe similar behaviour from the two. KDM performs slightly better, especially when also taking time complexity into account.

3.2 Results and analysis

In this section, we do not just provide our results but also compare with the sate-of-the-art methods that are relevant to our study as described in Section 3.1.4.

Table 2. Average recognition rate in % for GREC 2003 dataset.

Methods		ideal			scale		distort		Average
		set 1	set 2	set 3	set 1	set 2	set 1	set 2	
Bag-of-Relations	MBR	100	95	96	85	84	85	82	90
	RLM	100	100	100	100	92	100	88	97
Shape descriptors	$\mathcal{R}$-sign.	100	100	100	95	96	90	94	96
	ZM	100	100	100	100	98	95	92	97
	GFD	100	100	100	100	98	100	100	99
	SC	100	100	100	100	98	100	100	99
Pixel-based techniques	KDM	100	100	100	100	100	100	100	100
	SIHA	100	100	100	100	100	100	100	100

3.2.1 Test 1: GREC

In this test, every test image is used to find the closest model. The model is said to be recognised/found from which it produces the highest similarity score (see Eq. (10) in Section 2.4.2).

Based on this framework, Table 2 shows the average recognition rate. In this test, our method is not able to provide recognition rate of more than 97% from RLM that includes *extremities*. However, the performance is closer to a few shape descriptors such as $\mathcal{R}$-signature and ZM, and closer to GFD and SC. While comparing with the pixel-based approach that is specially designed for symbol recognition i.e., KDM, the BoR approach is less interesting.

On the whole, statistical shape representations are designed for isolated patterns, and can accommodate to noise. However, they difficultly accommodate with connected or composite symbols (FRESH dataset), which is the primary concern of this paper.

3.2.2 Test 2: FRESH

In this test, every test image (query) has been applied to retrieve similar images that may be isolated as well as the symbols in the composite form (as discussed in Section 3.1.1 and in Fig. 9). In our experiment, isolated symbols are only used for querying.

To make our study consistent, previous conclusions and test results are taken, and shown in Fig. 13. These reported results (including the quick comparison with the state-of-the-art methods) provide us a strong basis to compare with the proposed method. For the FRESH dataset, since it is composed of imbalanced ground-truths, we compute retrieval efficiency for 30 queries, and vary K from 1 to 10.

As reported in (Santosh, Lamiroy & Wendling, 2011), we observe the following.

1. RLM has been compared with common state-of-the-art spatial relation models as described in Section 3.1.4. Compared to MBR (taken from bench-marking relation models), RLM performs better with a substantial difference of more than 30%.
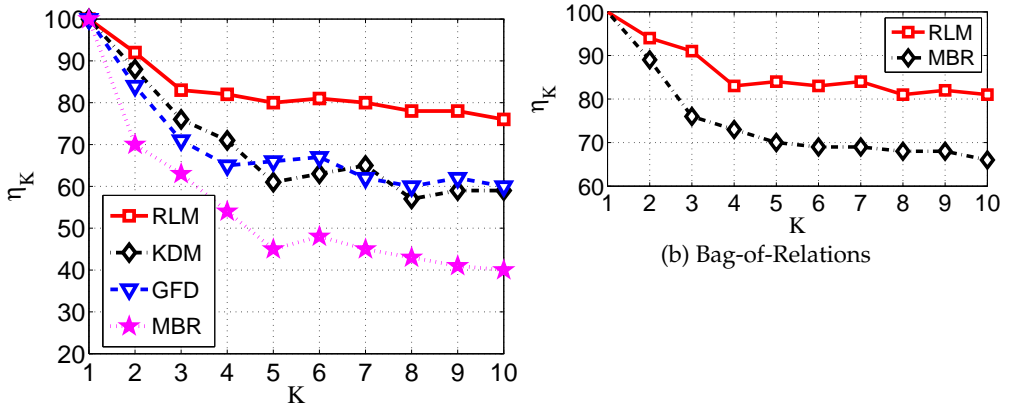
(a) Results in (Santosh, Lamiroy & Wendling, 2011)

Fig. 13. Average retrieval efficiency over requested list from 1 to 10 for FRESH dataset (a) state-of-the-art performance comparison reported in (Santosh, Lamiroy & Wendling, 2011) and (b) the performance of the proposed method (BORs).

2. Besides, a set of major state-of-the-art shape descriptors as described in Section 3.1.4 are taken, where GFD performs better among all but still under performs RLM by approximately more than 15% on average.

3. Furthermore, pixel-based approaches: SIHA and KDM, compared to RLM, perform similarly as GFD.

Fig. 13 (a) shows the performance comparison among all tested approaches.

In Fig. 13 (a), in Fig. 13 (b) we have employed RLM and MBR in our BoRs approach. Compared to K.C. *et al.*, 2011 (Santosh, Lamiroy & Wendling, 2011), BoRs provide a performance improvement by approximately 3–4% compared to RLM and 15–18% compared to MBR. Furthermore, our BoRs approach can be compared with the very recent works (Santosh, Lamiroy & Ropers, 2011; Santosh et al., 2014) where the significance of the shape descriptors on vocabulary has been studied. The comparison shows only a marginal difference (statistically similar results).

In Fig. 14, one can see whether the proposed method is able to retrieve significant known parts of the symbol from the composite form, when isolated symbols are used as queries. The example shows that the symbols with known significant parts of symbols (found anywhere in the image) can be retrieved. Thanks to the BORs, not all visual primitives from the database symbols are taken into account for relation alignment.

As mentioned in Section 1, at this point, it is important to comment on the previous approach (Santosh, Lamiroy & Wendling, 2011) that is relying on a ARG framework. It is limited to symbols having at least two vocabulary types since a single type (of any number of visual primitives) does not build a relational graph. As a consequence, isolated symbols (with a single vocabulary type) cannot be handled. For example, in the GREC dataset, a symbol
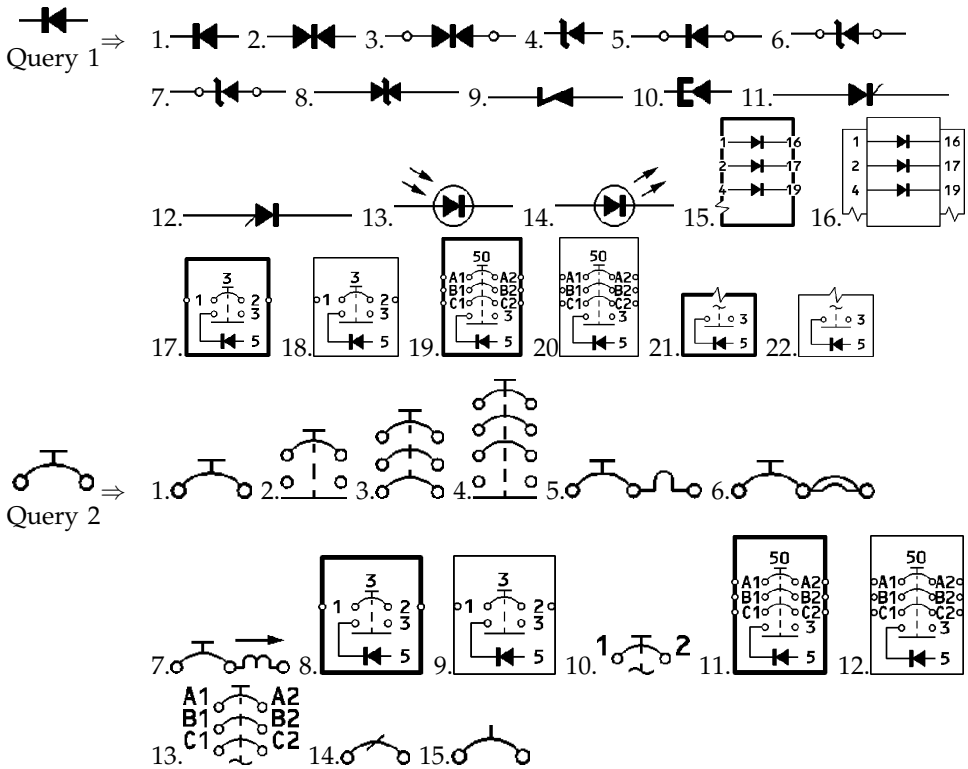
Fig. 14. Symbol retrieval (from FRESH dataset) using isolated symbol as queries.

can sometimes be composed of a collection of a single vocabulary type: four *corners* for any rectangle-shaped symbol.

3.2.3 Test 3: SESYD

Like the test 2, every model symbol is used as a query to retrieve the similar symbols from all database images. From each database image, multiple similar symbols are expected to be retrieved. Note that, since the ground-truth varies from one query symbol to another in each database image, we compute retrieval efficiency to evaluate the performance.

Fig. 15 shows retrieval efficiencies for both datasets: BoS and ED. For any query, the maximum number of similar symbols in the database image is less than 10. We however, compute retrieval efficiency over the short list that ranges from 1 to 10, expecting to retrieve all. In our experiments, not surprisingly, most of the retrieved symbols are matched with the ground-truths after top 5, and thus the retrieval efficiency curve goes up since it meets the criterion of recall (i.e., $N \leq K$ in Eq. (11) of Section 3.1.2). In this situation, one can achieve high recall. The similar phenomenon does not hold for test 2 since there are very few queries that have less than 10 similar symbols (ground-truths) in the database.
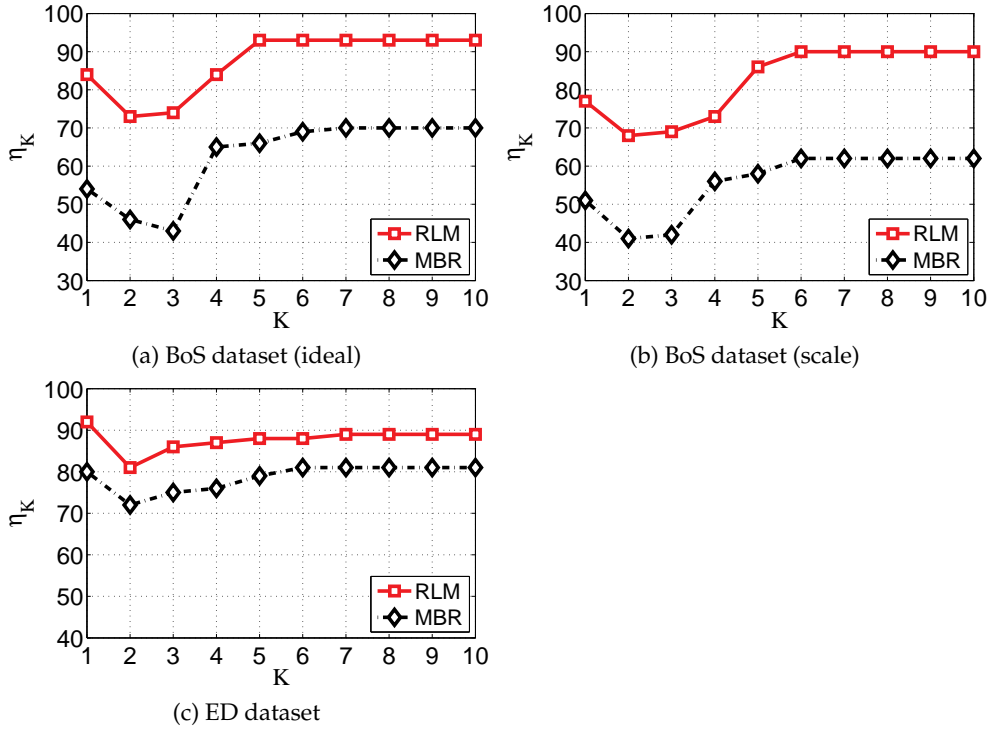
Fig. 15. Average retrieval efficiency over requested list ranging from 1 to 10 using SESYD datasets: (a) ideal symbols and (b) scaled symbols from BoS, and from (c) ED.

In Fig. 15, we observe that the performance reported in (c) is better than (a) and (b). This is primarily due to the selective or limited use of the models used for querying (see 10 models in Fig. 11). The model selection (from the ED dataset) has been made so that our vocabulary can appropriately be used. But, it does not hold the same for the BoS datasets. Furthermore, in the BoS dataset, more than 20% of the symbols are rejected since they are retrieved along with the parts of the closer symbols. This happens because a few visual primitives are missed from the symbols that are expected to be retrieved, and thus identical visual primitives can be matched (via relation alignment) from another symbol. In our test, the MBR model has been suffering more than RLM because MBR does not take relative distance (between the visual primitives) into account. This makes it easier for MBR to take visual primitives from other symbols in the database image. In contrast to isolated symbols (like in GREC 2003 dataset), the process (relation alignment) does not take visual primitives from other symbols.

Based on the appropriateness of our vocabulary, for visual understanding, Fig. 16 illustrates two examples (one per dataset) about symbol localisation and retrieval via RLM. These are taken as the best examples throughout the experiments. In this illustration, the colour of the rectangular box represents which query is used to extract which symbol from the image.
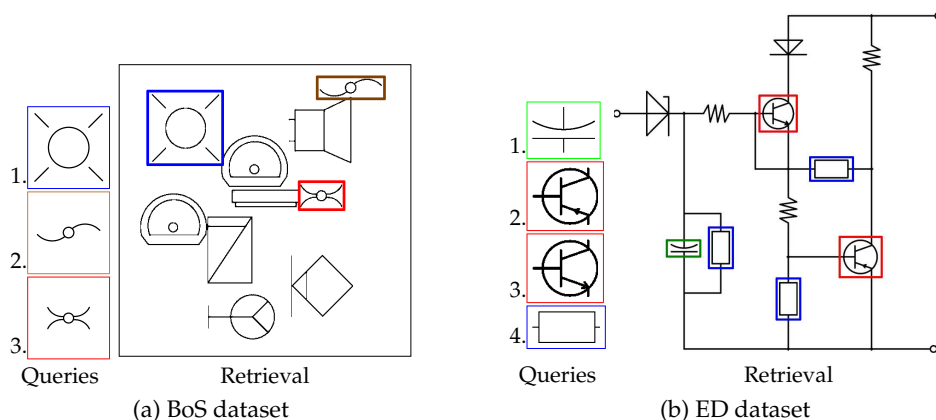
Fig. 16. Two examples (one per dataset) illustrating symbol retrieval by using (a) three different queries in BoS dataset and (b) four different queries in ED dataset.

In all tests, we did not compare benchmarking methods (like in other tests) into account since they require precise segmentation.

4. Extensions

Our results reported in Table 2, Figs. 13, and 15 however, do not significantly or fundamentally improve the state-of-the-art, nor extend the retrieval problem by localising a set of graphical elements. However, our method opens interesting new perspectives:

1. how well the symbols are described considering our vocabulary set; and
2. how flexible it is to select a query.

Item 1 will be discussed by proposing the best suited images (from wiring diagram) for our vocabulary unlike the SESYD BoS dataset. In *item 2*, flexibility refers to how flexible user inputs visual primitives as a query, instead of taking a complete symbol.

4.1 Vocabulary propriety

This section aims to show how appropriately our vocabulary can be used through a new experimental setup using wiring diagram. Having knowledge about graphical elements beforehand, we are able to detect significant or interested parts of wiring diagrams in test images. To handle this, other than the FRESH dataset, we have conducted another test:

1. A set of queries are manually selected from wiring diagrams, which we think are interesting, and by assuming that out vocabulary set can sufficiently describe their structures. Fig. 17(a) shows a set of queries.

(a) Query symbols

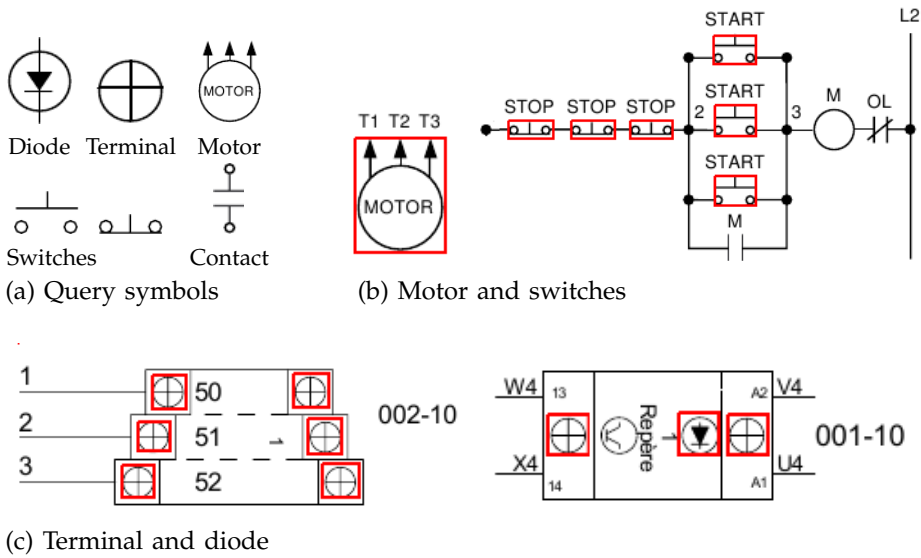(b) Motor and switches

(c) Terminal and diode

Fig. 17. Examples showing symbol localisation (in red) in the wiring diagrams by using the model symbols that are manually cropped.

2. Ten document images having wiring diagrams are taken for validation. In each document image, the extracted visual primitives are used to compute BoRs.

Fig. 17 shows a few examples of it using RLM, where localised symbols are bounded with a rectangular box in red. Fig. 17 shows queries in (a), motor and switches (open and close) are clearly localised in (b), and encircled diode (especially used in circuits like photo-sensor) and terminals are detected in (c). Even though the symbols are well localised, it is still interesting to see how well the ranking was. Fig. 18 shows the average retrieval efficiency over the requested list ranging from 1 to 10. In the reported result, we observe that the proposed method does not retrieve exactly similar symbols (defined as ground-truths) until top 6. After that, it retrieves all when RLM is employed. MBR has a similar behaviour (but with small difference).

Overall, this test has been made with small dataset based on our limited vocabulary set. A full framework of it can further be elaborated if one can extend to extract rich set of visual vocabulary.

4.2 User-friendly symbol localisation and retrieval

Following the visual illustrations in Figs. 14, 16 and 17, we are interested to use pair(s) of visual primitives as a query, that can be used to localise graphical elements in the database image. In other words, user takes a set of visual primitives which they are interesting. Such queries are graphically illustrated in Fig. 19. Since user provides their queries by selecting pair(s) of visual primitive, a user-friendly symbol retrieval test can be made.
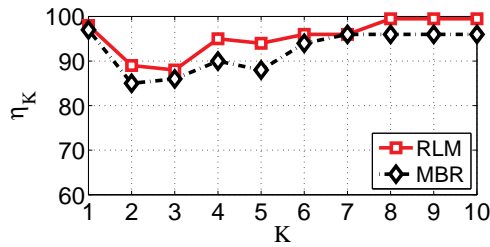
Fig. 18. Average retrieval efficiency over requested list ranging from 1 to 10 using wiring diagrams.

Query 1 *'Retrieve symbols with a thick inside a circle.'*

Based on the query description, symbols are retrieved. In this example, one can see no shape and size information about visual primitives has been taken into account.

Query 2 *'Retrieve rectangle-shaped symbols.'*

To illustrate it, we use a set of four *corners* facing to each other representing a *rectangle* and retrieve database symbols accordingly. In addition, our method also allows to count the number of rectangles included in the symbols. The idea is just to check whether separate corners are used for separate rectangles.

Query 3 *'Retrieve rectangle-shaped symbols containing a circle.'*

This query is employed since we aim to narrow down retrieval with respect to $\mathcal{Q}_2$. Additionally, in this example, an idea of no scaling effect has been addressed.

On the whole, our symbol localisation framework is invariant to

1. shape and size i.e., scaling, and
2. translation

as long as similar spatial organisation of visual primitives is found in the database symbol. Once again, the key concept is that user can take any spatial organisation of the visual primitives which they consider important. Using them as queries instead of using a whole symbol, can be used retrieve all related symbols from the database. Since we do not have ground-truths for those queries, we are not able to compute retrieval efficiency. But, the retrieval results are encouraging.

Additionally, this preliminary test is aimed to make connection with our previous work (Santosh et al., 2009), where we characterise the a set of symbols (with their counterparts) in terms a first-order-logic. In this work, we do the opposite, based on the user's interest.

5. Conclusions and future works

In this paper, we have presented a BoRs approach using the visual primitives that compose a symbol. Our approach uses topological relations to categorise visual primitives in the form
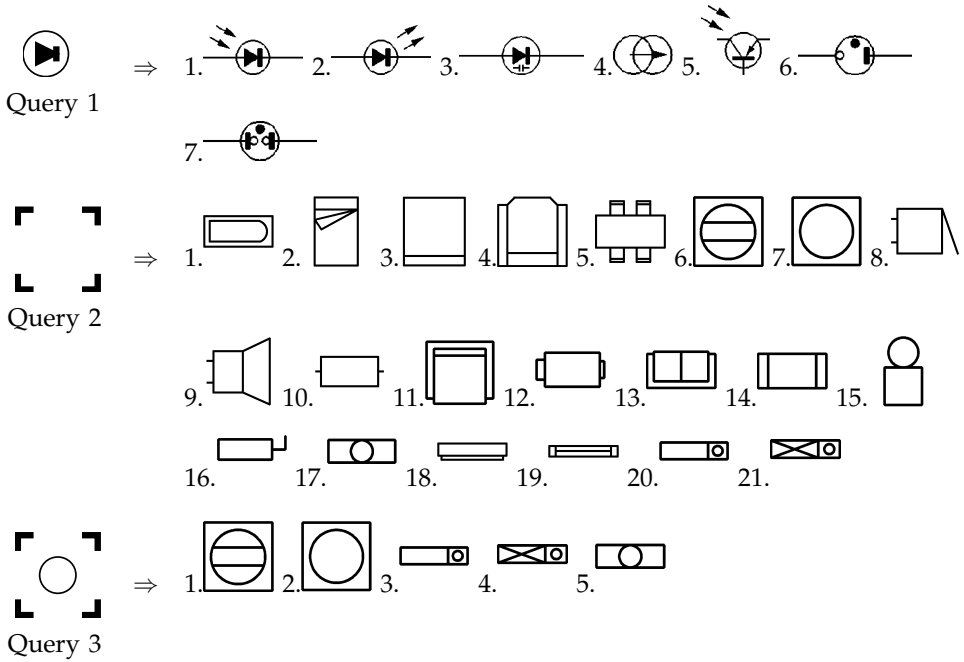
Fig. 19. Symbol retrieval using spatial organisation of visual primitives. For this, FRESH and GREC datasets are used.

of bags, and directional relations are computed and stored in addition to vocabulary type information. Such a topological guidance makes our method efficient in two different ways:

1. matching goes only to the relevant candidates i.e., it does not require all computed spatial relations; and

2. running time has been reduced i.e., rapid search is possible.

We have validated our method with a series experimental tests in a real-world configuration, including well known and publicly available datasets. Furthermore, we have also extended our tests in two different perspectives: vocabulary propriety (i.e., appropriate use of vocabulary) and user-friendly symbol localisation and retrieval. These preliminary tests can be taken as a strong basis for initiating symbol spotting, which is the contemporary issue in the domain (Rusiñol & Lladós, 2010). For this, considering our tests, we have observed that the vocabulary set has to be extended in accordance with the need. Further, it would be more appropriate to replace it with high-level vocabulary so that time complexity can be reduced.

After showing the possible extension of the vocabulary, it is interesting to use graph embedding techniques (bridging the gap between structural and statistical symbol recognition) and graph kernels, which are basically targeted to reduce the time complexity (Conte et al., 2013; Foggia & Vento, 2010; Luqman, 2012; Riesen & Bunke, 2010).

Additionally, working with deformed and degraded samples is in our plan, where we will take very recent work which is particularly focussed on geometric matching-based method (Nayef, 2012) into our comparative study.

Following our preliminary test that highlights user-friendly symbol localisation and retrieval (see Fig. 19 in Section 4.2), our another plan is to focus on how we can transform the idea of user inputs into first order logic (in the framework of natural language processing), not just limited to visual primitives.

6. Conflict of interest

None declared.

7. References

Ah-Soon, C. & Tombre, K. (2001). Architectural symbol recognition using a network of constraints, *Pattern Recognition Letters* **22**(2): 231–248.

Barrat, S. & 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. & Puzicha, J. (2002). Shape matching and object recognition using shape contexts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(4): 509–522.

Bodic, P. L., Locteau, H., Adam, S., Héroux, P., Lecourtier, Y. & Knippel, A. (2009). Symbol detection using region adjacency graphs and integer linear programming, *Proceedings of International Conference on Document Analysis and Recognition*, pp. 1320–1324.

Bunke, H. & Messmer, B. T. (1995). Efficient attributed graph matching and its application to image analysis, *in* C. Braccini, L. D. Floriani & G. Vernazza (eds), *Proceedings of International Conference on Image Analysis and Processing*, Vol. 974 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 45–55.

Bunke, H. & Messmer, B. T. (1997). Recent advances in graph matching, *International Journal of Pattern Recognition and Artificial Intelligence* **11**(01): 169–203.

Bunke, H. & Wang, P. (1997). *Handbook of Character Recognition and Document Image Analysis*, World Scientific.

Conte, D., Foggia, P., Sansone, C. & Vento, M. (2004). Thirty years of graph matching in pattern recognition, *International Journal of Pattern Recognition and Artificial Intelligence* **18**(3): 265–298.

Conte, D., Ramel, J.-Y., Sidere, N., Luqman, M. M., Gaüzère, B., Gibert, J., Brun, L. & Vento, M. (2013). A comparison of explicit and implicit graph embedding methods for pattern recognition, *in* W. G. Kropatsch, N. M. Artner, Y. Haxhimusa & X. Jiang (eds), *Proceedings of the IAPR Graph-Based Representations in Pattern Recognition*, Vol. 7877 of *Lecture Notes in Computer Science*, Springer, pp. 81–90.

Cordella, L. P. & Vento, M. (2000a). Symbol and shape recognition, *in* A. K. Chhabra & D. Dori (eds), *Graphics Recognition, Recent Advances*, Vol. 1941 of *Lecture Notes in Computer Science*, Springer, pp. 167–182.

Cordella, L. P. & 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. & 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.

Delalandre, M., Valveny, E., Pridmore, T. & Karatzas, D. (2010). Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems, *International Journal on Document Analysis and Recognition* **13**(3): 187–207.

Doermann, D. & Tombre, K. (2014). *Handbook of Document Image Processing and Recognition*, Springer-Verlag New York Incorporated.

Dosch, P. & Lladós, J. (2004). Vectorial signatures for symbol discrimination, *in* J. Lladós & Y.-B. Kwon (eds), *Graphics Recognition, Recent Advances and Perspectives*, Vol. 3088 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 154–165.

Dosch, P., Tombre, K., Ah-Soon, C. & Masini, G. (2000). A complete system for the analysis of architectural drawings, *International Journal on Document Analysis and Recognition* **3**(2): 102–116.

Egenhofer, M. & Franzosa, R. (1991). Point-set Topological Spatial Relations, *International Journal of Geographical Information Systems* **5**(2): 161–174.

Egenhofer, M. & Herring, J. R. (1991). *Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases*, University of Maine, Research Report.

Fletcher, L. & Kasturi, R. (1988). A robust algorithm for text string separation from mixed text/graphics images, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **10**(6): 910–918.

Foggia, P., Percannella, G. & Vento, M. (2014). Graph matching and learning in pattern recognition in the last 10 years, *International Journal of Pattern Recognition and Artificial Intelligence* **0**(0): 1450001.

Foggia, P. & Vento, M. (2010). Graph embedding for pattern recognition, *International Conference on Recognizing Patterns in Signals, Speech, Images and Videos*, Springer-Verlag, Berlin, Heidelberg, pp. 75–82.

GREC (2003). International symbol recognition contest at grec2003.
    URL: *http://www.cvc.uab.es/grec2003/SymRecContest/*

Güting, R. H. (1988). Geo-relational algebra: A model and query language for geometric database systems, *International Conference on Extending Database Technology: Advances in Database Technology*, pp. 506–527.

Kankanhalli, M. S., Mehtre, B. M. & Wu, J. K. (1995). Cluster-based color matching for image retrieval, *Pattern Recognition* **29**: 701–708.

Kara, L. B. & Stahovich, T. F. (2005). An image-based, trainable symbol recognizer for hand-drawn sketches, *Computers & Graphics* **29**(4): 501–517.

Kim, W.-Y. & Kim, Y.-S. (2000). A region-based shape descriptor using zernike moments, *Signal Processing: Image Communication* **16**(1-2): 95 – 102.

Lamiroy, B. & Guebbas, Y. (2010). Robust and precise circular arc detection, *in* J.-M. Ogier, L. Wenyin & J. Lladós (eds), *Graphics Recognition. Achievements, Challenges, and Evolution, 8th International Workshop, GREC 2009, La Rochelle, France, July 22-23, 2009. Selected Papers*, Vol. 6020 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 49–60.

Lee, W., Kara, L. B. & Stahovich, T. F. (2007). An efficient graph-based recognizer for hand-drawn symbols, *Computers & Graphics* **31**(4): 554–567.

Lladós, J., Martí, E. & 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. & Martí, E. (2002). Symbol Recognition: Current Advances and Perspectives, *in* D. Blostein & Y.-B. Kwon (eds), *GREC – Algorithms and Applications*, Vol. 2390 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 104–127.

Lladós, J. & Sánchez, G. (2004). Graph matching versus graph parsing in graphics recognition – a combined approach, *International Journal of Pattern Recognition and Artificial Intelligence* **18**(03): 455–473.

Luqman, M. M. (2012). *Fuzzy Multilevel Graph Embedding for Recognition, Indexing and Retrieval of Graphic Document Images*, PhD thesis, FranÃ§ois Rabelais University of Tours France and Autonoma University of Barcelona Spain.

Messmer, B. T. & Bunke, H. (2000). Efficient subgraph isomorphism detection: A decomposition approach, *IEEE Trans. Knowl. Data Eng.* **12**(2): 307–323.

Miyajima, K. & Ralescu, A. (1994). Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations, *Fuzzy Sets and Systems* **2**(65): 225–236.

Nagy, G. (2000). Twenty years of document image analysis in pami, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1): 38–62.

Nayef, N. (2012). *Geomatric-Based Symbol Spotting and Retrieval in Technical Line Drawings*, PhD thesis, University of Kaiserslautern, Germany.

Papadias, D. & Theodoridis, Y. (1997). Spatial relations, minimum bounding rectangles, and spatial data structures, *International Journal of Geographical Information Science* **11**(2): 111–138.

Pullar, D. & Egenhofer, M. (1988). Towards formal definitions of topological relations among spatial objects, *In: D. Marble, editor, The Third International Symposium on Spatial Data Handling*, pp. 225–242.

Qureshi, R. J., Ramel, J.-Y., Barret, D. & Cardot, H. (2008). Spotting symbols in line drawing images using graph representations, *in* W. Liu, J. Lladós & J.-M. Ogier (eds), *Proceedings of IAPR International Workshop on Graphics Recognition*, Vol. 5046 of *Lecture Notes in Computer Science*, Springer, pp. 91–103.

Rendek, J., Masini, G., Dosch, P. & Tombre, K. (2004). The search for genericity in graphics recognition applications: Design issues of the qgar software system, *in* S. Marinai & A. Dengel (eds), *Proceedings of International Workshop on Document Analysis Systems*, Vol. 3163 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 366–377.

Riesen, K. & Bunke, H. (2010). *Graph Classification and Clustering Based on Vector Space Embedding*, World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Robles-Kelly, A. & Hancock, E. R. (2004). String edit distance, random walks and graph matching, *International Journal of Pattern Recognition and Artificial Intelligence* **18**(03): 315–327.

Rusiñol, M. & Lladós, J. (2006). Symbol spotting in technical drawings using vectorial signatures, *in* W. Liu & J. Lladós (eds), *Graphics Recognition. Ten Years Review and Future Perspectives*, Vol. 3926 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 35–46.

Rusiñol, M. & Lladós, J. (2010). *Symbol spotting in digital libraries: Focused retrieval over graphic-rich document collections*, Springer Verlag.

Salmon, J. P., Wendling, L. & 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. & Ropers, J.-P. (2009). Inductive logic programming for symbol recognition, *Proceedings of International Conference on Document Analysis and Recognition*, pp. 1330–1334.

Santosh, K. C., Lamiroy, B. & Ropers, J.-P. (2011). Spatio-structural symbol description with statistical feature add-on, *Proceedings of IAPR International Workshop on Graphics Recognition*.

Santosh, K. C., Lamiroy, B. & Wendling, L. (2011). Symbol recognition using spatial relations, *Pattern Recognition Letters* **33**(3): 331–341.

Santosh, K. C., Lamiroy, B. & Wendling, L. (2014). Integrating vocabulary clustering with spatial relations for symbol recognition, *International Journal on Document Analysis and Recognition* **17**(1): 61–78.

Santosh, K. C., Wendling, L. & Lamiroy, B. (2013). Relation bag-of-features for symbol retrieval, *Proceedings of International Conference on Document Analysis and Recognition*, IEEE, pp. 768–772.

Tabbone, S., Wendling, L. & Salmon, J.-P. (2006). A new shape descriptor defined on the radon transform, *Computer Vision and Image Understanding* **102**(1): 42–51.

Tabbone, S., Wendling, L. & Zuwala, D. (2004). A hybrid approach to detect graphical symbols in documents, *in* S. Marinai & A. Dengel (eds), *Proceedings of International Workshop on Document Analysis Systems*, Lecture Notes in Computer Science, Springer, pp. 342–353.

Terrades, O. R., Valveny, E. & Tabbone, S. (2008). On the combination of ridgelets descriptors for symbol recognition, *in* L. Wenyin, J. Lladós & J.-M. Ogier (eds), *Graphics Recognition. Recent Advances and New Opportunities*, Vol. 5046 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 40–50.

Terrades, O. R., Valveny, E. & 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., Ah-Soon, C., Dosch, P., Habed, A. & Masini, G. (1998). Stable, robust and off-the-shelf methods for graphics recognition, *Proceedings of International Conference*

on Pattern Recognition **1**: 406.

Tooley, M. & Wyatt, D. (2008). *Aircraft electrical and electronic systems: principles, operation and maintenance*, Aircraft engineering principles and practice, Butterworth-Heinemann.

Valveny, E. & Martí, E. (2003). A model for image generation and symbol recognition through the deformation of lineal shapes, *Pattern Recognition Letters* **24**(15): 2857–2867.

Visani, M., Bertet, K. & Ogier, J.-M. (2011). Navigala: An original symbol classifier based on navigation through a galois lattice, *International Journal of Pattern Recognition and Artificial Intelligence* **25**(04): 449–473.

Wang, X. & Keller, J. (1999). Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches, *Fuzzy Sets and Systems* **101**: 5–20.

Wenyin, L., Zhang, W. & 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. & 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.

Yoon, S., Lee, Y., Kim, G. & Choi, Y. (2001). New paradigm for segmentation and recognition of handwritten numeral string, *Proceedings of International Conference on Document Analysis and Recognition*, pp. 205–209.

Yuen, P. C., Feng, G.-C. & 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. & Lu, G. (2002). Shape-based image retrieval using generic fourier descriptor, *Signal Processing: Image Communication* **17**: 825–848.

Zhang, D. & Lu, G. (2004). Review of shape representation and description techniques, *Pattern Recognition* **37**(1): 1–19.

Zhang, W., Wenyin, L. & Zhang, K. (2006). Symbol recognition with kernel density matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12): 2020–2024.

Authors' biography

K.C. Santosh is currently a research fellow at the Lister Hill National Center for Biomedical Communications, an R&D division of the U.S. National Library of Medicine (NLM), National Institutes of Health (NIH), USA. Before this, K.C. worked as a postdoctoral research scientist at the LORIA research centre, Université de Lorraine and ITESOFT, France. K.C. earned his Ph.D. in Computer Science from INRIA Nancy Grand Est, Université de Lorraine, France in 2011 under the FP-6 European project (and INRIA-CORDI research grant), Master of Science diploma in Information Technology by research and thesis from the school of Information and Communication Technology, SIIT, Thammasat University, Thailand in 2007 under the Asian Development Bank Japan Scholarship Program (ADB-JSP), and Bachelor of Engineering in Electronics and Communication from Pokhara University, Nepal, in 2003, under a merit scholarship.

His research interests include document image analysis, document information content exploitation, biometrics (such as face) and (bio)medical image analysis. He has served (serves) on the program committee of conferences such as ICDAR 2013, ICPR 2014, Bioimaging 2015, VISAPP 2015, ICPRAM 2015. He also serves on the editorial board of Journal of Pattern Recognition Research.

Laurent Wendling was born in Lunéville, France, in 1969. He received the Ph.D. degree in computer science from the University of Paul Sabatier, Toulouse, France, in 1997. He received the "HDR" degree in 2006. From 1993 to 1999, he was with the IRIT in the field of pattern recognition. From 1999 to 2009, he was an Assistant Professor at the ESIAL Nancy, and a member of LORIA in the field of symbol recognition. His current research topics are spatial relation, feature selection, and image segmentation. He is currently a full professor at the Paris Descartes University, in the field of computer science. He is also the group leader of the SIP team, LIPADE.

Bart Lamiroy is a permanent faculty member at the Université de Lorraine, in Nancy, France, and member of the associated LORIA research lab since 2000. He was a visiting scientist at Lehigh University from January 2010 to July 2011. He has a broad experience in Machine Perception. Over the years, his research topics have ranged from Content Based Image Retrieval over Visual Servoing to Document Analysis.

From 2007 to 2009, he was head of the Computer Science and IT

Department at the École des Mines de Nancy, France, and has been a permanent faculty member there since 2000. Before that he was a research contractor at INRIA, after having obtained his Ph.D. in computer vision at the Institut National Polytechnique de Grenoble, France in 1998. He received his bachelor's degree in applied mathematics in 1993. He also serves on the International Association for Pattern Recognition TC-10 Committee as Dataset Curator, and on the Publicity and Publications Committee.