

Relation Bag-of-Features for Symbol Retrieval

Santosh K.C., Laurent Wendling, Bart Lamiroy

► **To cite this version:**

Santosh K.C., Laurent Wendling, Bart Lamiroy. Relation Bag-of-Features for Symbol Retrieval. IC-DAR - International Conference on Document Analysis and Recognition - 2013, Aug 2013, Washington DC, United States. hal-00823960

HAL Id: hal-00823960

<https://hal.inria.fr/hal-00823960>

Submitted on 20 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.

Relation Bag-of-Features for Symbol Retrieval

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

[†]LORIA – Université de Lorraine

BP - 239, Loria Campus Scientifique, 54506 Nancy, FRANCE

[‡]LIPADE – Université Paris Descartes

75270 Paris, FRANCE

Email: {santosh.kc, bart.lamiroy}@loria.fr, laurent.wendling@parisdescartes.fr

Abstract—In this paper, we address a new scheme for symbol retrieval based on relation bag-of-features (BOFs) which are computed between the extracted visual primitives. Our feature consists of pairwise spatial relations from all possible combinations of individual visual primitives. The key characteristic of the overall process is to use topological information to guide directional relations. Consequently, directional relation matching takes place only with those candidates having similar topological configurations. A comprehensive study is made by using two different datasets. Experimental tests provide interesting results by establishing user-friendly symbol retrieval application.

I. INTRODUCTION

Document analysis or processing is mainly related to both text and graphic separation, localisation and recognition. According to [1], 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.

Graphics recognition has an extremely rich state-of-the-art literature in symbol recognition and localisation since the 70's. However, most methods are particularly suited for isolated line symbols, not for composed symbols connected to a complex environment [2], [3]. Considering the problem of symbol localisation in real documents, composed of individual parts and constrained by spatial relations for instance, global signal based descriptors [4] cannot be applied since they are, unfortunately, primarily designed for applications where line symbols are isolated [5]. As a consequence, one needs to be able to extract visual primitives like *points*, *lines* and *arcs* and formalise the possible links that exist between them. These primitives are then used in structural descriptors like attributed relational graphs (ARG) [6], region adjacency graphs (RAG) [7], as well as deformable templates [8]. In this context, the variability of the size of the underlying graph structures leads to computational complexity in matching. Structural approaches however, provide a powerful relational representation, conveying how parts are connected to each other, and are usually considered to be preserving generality and extensibility.

In this paper, we address some of the shortcomings of previous work [9] which combined structural approaches and spatial relations and extend its scope so that it can be used in a context of flexible querying, retrieval and localisation of symbols. This previous work [9] decomposes a symbol in

specifically detected elements and uses them as a basis for expressing and computing spatial relations. These elements are further grouped together by type to avoid the NP-hardness of the underlying graph matching problem. However, it needs at least two different types of visual components to compute the required spatial relations, and is therefore unsuited for symbols having only a single vocabulary type (e.g. *corners* only) – no matter how many visual primitives it contains. The extension and improvements presented in this paper address this by computing all possible relations that exist between individual visual primitives, rather than relying on classes of primitives. Execution time is reduced by the use of *Bag-of-Features* (BOF) indexing.

The rest of the paper is organised as follows: in Section II, we explain our method mainly discussing relation BOFs; in Section III our recognition technique is explained; full experimental results are reported in Section IV. The paper is concluded in Section V.

II. RELATION BAG-OF-FEATURES

Given a predefined application context, we define what is semantically significant (*circle* and *corner* types, for instance), and extract these primitives from the image. The main idea developed in this section is the construction of "relation BOFs", where all possible pairs of visual primitives pairs have their topological and directional relations computed [9] and are indexed with respect to these relation values. In other words, 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 information that 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 that integrating 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 features of the visual primitives that compose the symbol, directional relations are computed and stored in each particular bag, as shown below.

A. Visual vocabulary

We define a set of well controlled visual primitives as a *vocabulary* using standard image analysis operators. Our current vocabulary set consists of *thick*, *circle* and *corner* components. Fig. 1 shows a few examples. The choice of

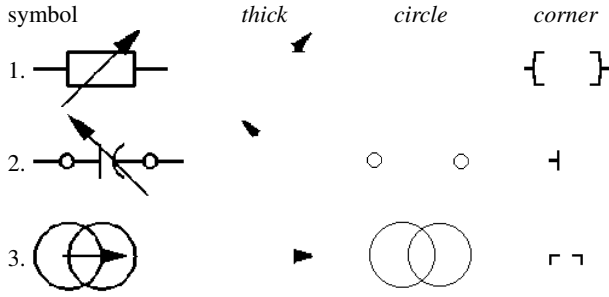


Fig. 1. Visual vocabulary from a few corresponding symbols.

the vocabulary depends on the application context, and can be of any kind or type, related to what is visually pertinent in that particular context. Ours is related to electrical symbols. In what follows, we shall refer to the set of vocabulary types as, $\{\mathbb{T}_{thick}, \mathbb{T}_{circle}, \mathbb{T}_{corner}\}$.

B. 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 $1 \leq t \leq 3$). For any vocabulary type \mathbb{T}_t , there are m_t visual primitives,

$$\mathbb{T}_t = \{\varphi_i^t\}, i = [1, \dots, m_t] \text{ and } p = \sum_t m_t.$$

Any pair of primitives (φ_1, φ_2) , as illustrated in Fig. 2, 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).

We shall refer to this set of relations as $\{\mathcal{C}_k\}, k = [1, \dots, 6]$.

To obtain the topological relation $\mathcal{T}(\varphi_1, \varphi_2)$, we use the 9-intersection model [10], [11] relative to the boundaries (∂^*), interiors ($*^o$) and exteriors ($*^-$) of φ_1 and φ_2 . Their definitions use basic set operations like $=, \neq, \subseteq$ and \cap [10]. For example,

- $equal(\varphi_1, \varphi_2) := \text{points}(\varphi_1) = \text{points}(\varphi_2)$;
- $disconnected(\varphi_1, \varphi_2) := \text{points}(\varphi_1) \neq \text{points}(\varphi_2)$ or $\text{points}(\varphi_1) \cap \text{points}(\varphi_2) = \emptyset$; and
- $inside(\varphi_1, \varphi_2) := \text{points}(\varphi_1) \subseteq \text{points}(\varphi_2)$.

Taking a few examples from Fig. 1, we obtain:

- 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* . The remaining *corners* are *disconnected* to the *thick*. In a similar manner, there exist *externally connected* and *disconnected* intra-corner topological configurations.

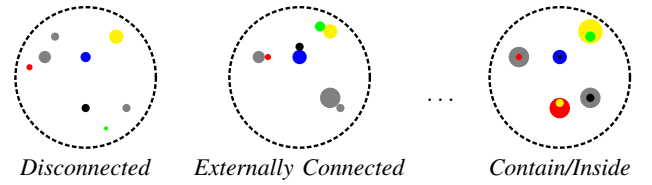


Fig. 2. Our relation BOFs model – a symbolic example. Each item in every bag, represents a visual primitive and its colour represents its type.

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* that are *externally connected*. As in item 1, we have two different bags: *disconnected* and *externally connected*.

C. Spatial reasoning

We have categorised visual primitives by using their topological relations their vocabulary type information i.e., $(\varphi_i, \varphi_j) \mapsto ((\mathbb{T}_t, \mathbb{T}_{t'}), \mathcal{T}(\varphi_i, \varphi_j))$. We now compute directional relations \mathcal{D} between them,

Pairwise spatial relations are often expressed by using one of the objects as reference. For example, φ_1 is to the *right* of φ_2 : $right(\varphi_1, \varphi_2)$, where φ_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 point \mathbb{R} from each pair and compute directional relations with respect to it. It can thus avoid potential ambiguity.

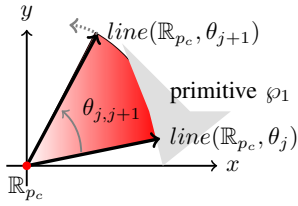
Unique reference point set (\mathbb{R}). It is either the common portion of two neighbouring sides in the case of *disconnected* primitives or the intersection in the case of *overlapping*, *equal* or otherwise *connected* primitives. For computational reasons, the visual primitives are encased within their minimum boundary rectangles (MBR).

Directional relations. To compute directional relations between the primitives in all categories, we use the radial line model (RLM) presented in [9]. It should be noted that other spatial relation models could be used as well.

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,

$$\mathcal{H}_{\varphi_1, \mathbb{R}_{p_c}} = \left[\frac{sector(\mathbb{R}_{p_c}, \theta_{j,j+1}) \cap \text{points}(\varphi_1)}{\text{points}(\varphi_1)} \right]_{j=0, \dots, s-1}.$$

Fig. 3 shows an illustration of this process. It shows how a sector can be built using two consecutive lines and how



(a) Sector made by two consecutive radial lines in \mathbb{R}^2 space. (b) A single histogram value in a defined sector.

Fig. 3. An illustration of how histogram is computed in a sector made by two consecutive radial lines at θ_j and θ_{j+1} .

the histogram can be computed in that sector. At this point, it is important to notice that value of $\theta_{j,j+1}$ determines the precision (or accuracy) of the histogram. The lower $\theta_{j,j+1}$, the better the performance [9]. The directional relation \mathcal{D} between φ_1 and φ_2 is defined by combining the histograms corresponding to both elements as

$$\mathcal{D}(\varphi_1, \varphi_2) = \{\mathcal{H}_{\varphi_1, \mathbb{R}_{p_c}}, \mathcal{H}_{\varphi_2, \mathbb{R}_{p_c}}\}.$$

An example is given in Fig. 4.

III. RECOGNITION

Now that we have defined type, topology and directional signatures, we can describe our recognition method. Symbol recognition is based on relation matching between corresponding pairs of visual primitives. However, in order to reduce combinatorial complexity we only consider configurations where matching candidate pairs share the exact same values for their vocabulary types and topological relations.

A. Relation matching

Consider a symbol \mathcal{S} having a set of visual primitives $\{\varphi_i^t\}$ (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 $\{\mathcal{C}_k\}$ and can be attributed a directional relation signature $\mathcal{D}_k(\varphi_i^t, \varphi_{i'}^{t'})$.

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

- 1) For each query pair $(\varphi_i^t, \varphi_{i'}^{t'})$ of \mathcal{S}^q , we select the bag of database pairs $(\varphi_{\tilde{i}}^t, \varphi_{\tilde{i}'}^{t'})$ from \mathcal{S}^d belonging to the same topological relation category \mathcal{C}_k .
- 2) Within this category, we match query and database pairs by computing the distance between their directional relation signature $\mathcal{D}(\cdot, \cdot)$.

Matching only takes place between pairs sharing the exact same vocabulary types. i.e., between $(\varphi_i^t, \varphi_{i'}^{t'})$ and $(\varphi_{\tilde{i}}^t, \varphi_{\tilde{i}'}^{t'})$ whose vocabulary types \mathbb{T}_t and \mathbb{T}_t' are identical i.e.,

$$\delta \left(\mathcal{D}_k^q(\varphi_i^t, \varphi_{i'}^{t'}), \mathcal{D}_k^d(\varphi_{\tilde{i}}^t, \varphi_{\tilde{i}'}^{t'}) \right),$$

where $\delta(a, b) = ||a - b||$. The closest matching distance pairs will be $\arg \min_{\tilde{i}\tilde{i}'} \delta(\cdot, \cdot)$ and for all query pairs $pair_{ii'}$ in a particular category k , distance can be aggregated as,

$$\Delta_k(\mathcal{S}^q, \mathcal{S}^d) = \sum_{ii'} \min_{\tilde{i}\tilde{i}'} \delta \left(\mathcal{D}_k^q(\varphi_i^t, \varphi_{i'}^{t'}), \mathcal{D}_k^d(\varphi_{\tilde{i}}^t, \varphi_{\tilde{i}'}^{t'}) \right).$$

Finally, taking into account all pre-defined categories, the total distance between two symbols becomes

$$\text{Dist.}(\mathcal{S}^q, \mathcal{S}^d) = \sum_k \Delta_k(\cdot, \cdot).$$

B. Recognition and retrieval

The $\text{Dist.}(\cdot, \cdot)$ 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 $\text{Dist.}(\mathcal{S}^q, \mathcal{S}^d)$, $d \in \{1, \dots, \mathbb{D}\}$ with $\overline{\text{Dist.}}(\cdot, \cdot) = \frac{\text{Dist.}(\cdot, \cdot) - \text{Dist.}^{\min.}(\cdot, \cdot)}{\text{Dist.}^{\max.}(\cdot, \cdot) - \text{Dist.}^{\min.}(\cdot, \cdot)}$. We then express similarity as, $\text{Similarity}(\mathcal{S}^q, \mathcal{S}^d)$

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

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).

IV. EXPERIMENT

A. Datasets and evaluation metrics

Datasets. To validate the proposed relation BOFs method, we have primarily tested 1) FRESH dataset [9] and 2) GREC dataset [12]. The latter dataset aims to provide the generality of the approach. FRESH dataset is composed of roughly 500 different known symbols, some of which come from [13]. It shows that symbols may either be very similar in shape – and only differ by slight details – or 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. In case of GREC dataset, there are altogether 50 different model symbols. Those symbols are grouped into three sets, containing 5, 20 and 50 model symbols. Among many, ideal, scaling and distortion models are used.

Evaluation metrics. For FRESH dataset, since the number of similar symbols (in ground-truths), may vary largely from one query to another, we use retrieval efficiency [14] as a measure for retrieval quality. For a chosen query and for a fixed number of K returned symbols, it can be expressed as,

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

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. In case of GREC dataset, we employ recognition rate since dataset is fully balanced and composed of labelled ground-truths.

B. Results and analysis

Since our approach is an extension of the work [9], we can admit some of the conclusions in the latter, and focus only on some essential performance parts without needing to proceed with comparison to the complete state-of-the-art, already presented. In this framework, we primarily focus on

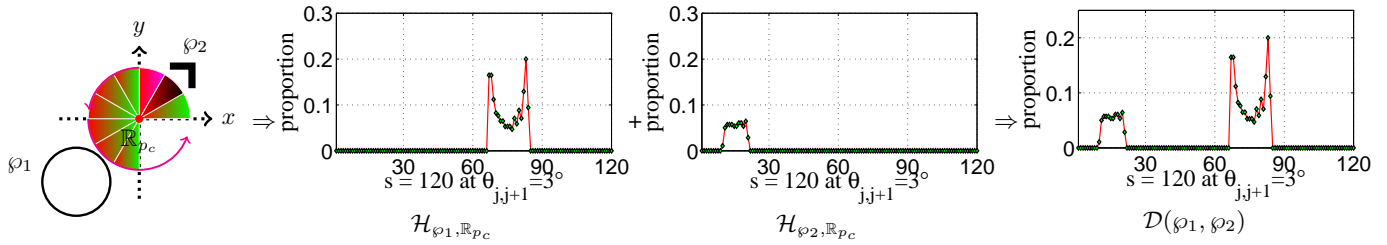


Fig. 4. A disconnected pair of φ_1^{circle} and φ_2^{corner} including unique reference point set \mathbb{R} and directional relational histograms using RLM with respect to \mathbb{R}_{p_c} . RLM is applied for both visual primitives and combined together.

the FRESH dataset [13]. For evaluation, we have used retrieval efficiency, as described in Section IV-A. These efficiency values have been computed for values of $K = 1$ to 10, over 30 queries.

We perform a series of experimental tests by using the radial line model¹ [9] and projection model [15]. Our aim is to attest the BOFs approach where relations are taken as candidate features. Figure 5 shows retrieval efficiency for FRESH dataset over requested list: 1 to 10 and recognition rates for GREC dataset. Compared to [9], for FRESH dataset, relation BOFs possess similar behaviour but with a small improvement (3–4%) from both relation models. In case of GREC dataset, we are not able to provide more than 97%. At this point, it is important to notice that [9] was inoperant on the GREC dataset when a symbol was composed of a collection of only a single visual primitive type: four *corners* for any rectangle-shaped symbol, for instance.

Considering matching time complexity issue, our approach avoids to process the whole database. This is due to vocabulary type information as well as categorisation principle used in the BOFs. As a consequence, an running time is reduced by more than a factor of 4 in average with respect to when no vocabulary categorisation is used.

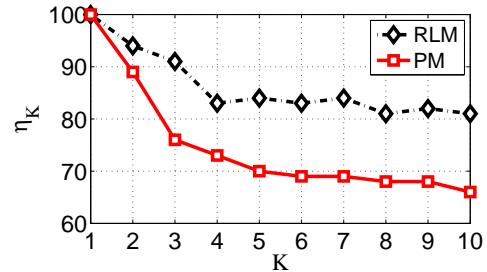
C. Extensions

Our results reported in Fig. 5 however, simply do not project real advantages in the context of retrieval problem. To forecast it (while maintaining an appropriate use), we explicitly analyse and extend the approach in two different perspectives: 1) whether the proposed method is able to retrieve significant known parts of the symbol from the composite form; and 2) flexibility in selecting a query².

Both perspectives are illustrated in Fig. 6. In case of query Q_1 , we provide an isolated symbol as a query and their corresponding retrieved symbols from the database (cf. Fig. 6). The example shows that the symbols with known significant parts of symbols (found anywhere in the image) can be retrieved. Based on retrieval performance using query Q_1 , we are interested to use spatial organisation of visual primitives as query, that can be used for localising graphical elements. It eventually provides a user-friendly symbol retrieval. Let us explain it in more details with the help of the following queries

¹In case of radial line model, we do not consider the influence of different angular steps that represents the trade-off between the optimal choice of resolution – and thus precision of spatio-structural information capture – and time/space requirements. As in the original work [9], we use 3° angular step.

²Flexibility – it refers to how flexible user inputs visual primitives and their spatial organisation as a query, instead of taking a complete symbol.



(a) FRESH dataset (retrieval efficiency η_K)

	ideal			scale		distort	
	set 1	set 2	set 3	set 1	set 2	set 1	set 2
PM	100	95	96	84	84	88	82
RLM	100	100	100	100	92	100	88

(b) GREC dataset (recognition rate %)

Index

RLM: radial line model and
PM: projection model.

Fig. 5. Performance of the relation BOFs: (a) Average retrieval efficiency over requested list: 1 to 10 for FRESH dataset and (b) Recognition rate in % for GREC2003 dataset.

Q_2 to Q_4 where user selects pairs of visual primitives which they think are important, aiming to retrieve significant known parts of the symbols i.e., graphical elements in the image. It eventually provides a user-friendly symbol retrieval application in connection with [16]:

Q_2 . ‘Retrieve symbols with a thick inside a circle.’

Based on the query description, symbols are retrieved. In this example, we address spatial relations between the visual primitives without considering shape and size information.

Q_3 . ‘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. The idea is just to check whether separate corners are used for separate rectangles.

Q_4 . ‘Retrieve rectangle-shaped symbols containing a circle.’

This query is employed since we aim to narrow down retrieval with respect to Q_3 . In addition, in this illustration, it addresses an idea of no scaling effect in retrieval.

On the whole, our symbol retrieval attests the fact that there exist no shape and size i.e., scaling, and translation effect in retrieval as long as similar spatial organisation of visual primitives can be found in the database symbol. In contrast to common state-of-the-art approaches (cf. Section I), our method provides an interesting symbol retrieval application.

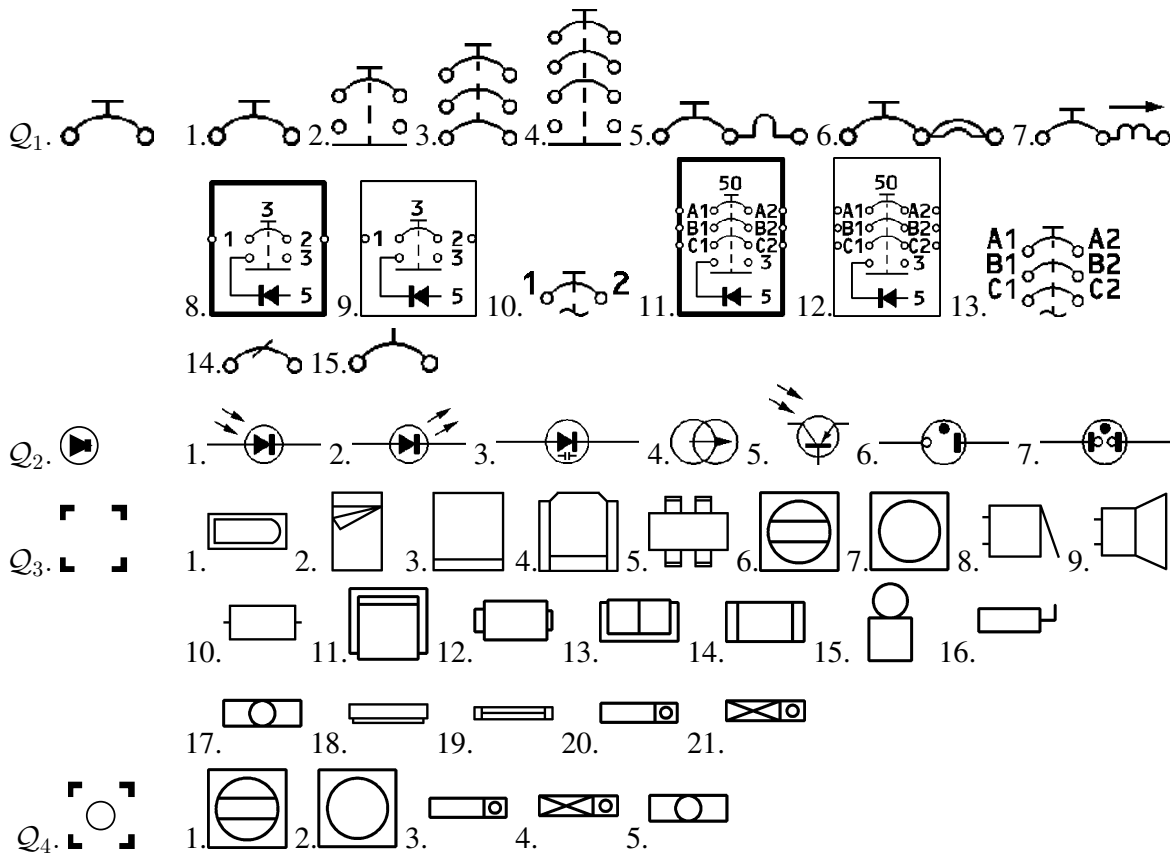


Fig. 6. Symbol retrieval using isolated symbol as a query Q_1 and spatial organisation of visual primitives $Q_2 : Q_4$. The first two queries are related with FRESH dataset and the remaining ones are related with the GREC dataset. The use of boolean relations (PM model) allows to produce qualitative retrieval.

V. CONCLUSIONS

In this paper, we present relation BOFs approach using the visual primitives that compose a symbol. Our approach uses topological relations to categorise visual primitives in the form of bags where 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 drastically reduced i.e., faster search is possible. We have validated our method with a series experimental tests in the real-world industrial problem. We have also demonstrated the extension of the method by providing a user-friendly symbol retrieval application including the screen-shot of the symbol spotting perspectives in the domain [17].

REFERENCES

- [1] G. Nagy, "Twenty years of document image analysis in pami," *PAMI*, vol. 22, no. 1, pp. 38–62, 2000.
- [2] L. P. Cordella and M. Vento, "Symbol recognition in documents: a collection of techniques?" *IJDAR*, vol. 3, no. 2, pp. 73–88, 2000.
- [3] J. Lladós, E. Valveny, G. Sánchez, and E. Martí, "Symbol Recognition: Current Advances and Perspectives," in *GREC – Algorithms and Applications*, ser. LNCS, vol. 2390. Springer-Verlag, 2002, pp. 104–127.
- [4] D. Zhang and G. Lu, "Review of shape representation and description techniques," *PR*, vol. 37, no. 1, pp. 1–19, 2004.
- [5] L. P. Cordella and M. Vento, "Symbol and shape recognition," in *Graphics Recognition, Recent Advances*, ser. LNCS, vol. 1941. Springer, 2000, pp. 167–182.
- [6] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *IJPRAI*, vol. 18, no. 3, pp. 265–298, 2004.
- [7] J. Lladós, E. Martí, and J. J. Villanueva, "Symbol recognition by error-tolerant subgraph matching between region adjacency graphs," *PAMI*, vol. 23, no. 10, pp. 1137–1143, 2001.
- [8] E. Valveny and E. Martí, "A model for image generation and symbol recognition through the deformation of lineal shapes," *PR Letters*, vol. 24, no. 15, pp. 2857–2867, 2003.
- [9] K. C. Santosh, B. Lamiroy, and L. Wendling, "Symbol recognition using spatial relations," *PR Letters*, vol. 33, no. 3, pp. 331–341, 2011.
- [10] R. H. Güting, "Geo-relational algebra: A model and query language for geometric database systems," in *Intl. Conf. on Extending Database Techno.: Advances in Database Techno.*, 1988, pp. 506–527.
- [11] M. Egenhofer and J. R. Herring, *Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases*. University of Maine, Research Report, 1991.
- [12] GREC, "International symbol recognition contest at grec2003," 2003. [Online]. Available: <http://www.cvc.uab.es/grec2003/SymRecContest/>
- [13] M. Tooley and D. Wyatt, *Aircraft electrical and electronic systems: principles, operation and maintenance*, ser. Aircraft engineering principles and practice. Butterworth-Heinemann, 2008.
- [14] M. S. Kankanhalli, B. M. Mehre, and J. K. Wu, "Cluster-based color matching for image retrieval," *PR*, vol. 29, pp. 701–708, 1995.
- [15] D. Papadias and Y. Theodoridis, "Spatial relations, minimum bounding rectangles, and spatial data structures," *IJGIS*, vol. 11, no. 2, pp. 111–138, 1997.
- [16] K. C. Santosh, B. Lamiroy, and J. P. Ropers, "Inductive Logic Programming for Symbol Recognition," in *ICDAR*, 2009, pp. 1330 – 1334.
- [17] M. Rusiñol and J. Lladós, *Symbol spotting in digital libraries: Focused retrieval over graphic-rich document collections*. Springer Verlag, 2010.