

Browsing graphics without prior knowledge

Daniel Zuwala, Jan Rendek

▶ **To cite this version:**

Daniel Zuwala, Jan Rendek. Browsing graphics without prior knowledge. International Conference on Pattern Recognition 2006 - ICPR 2006, Aug 2006, Hong Kong (China). inria-00078450

HAL Id: inria-00078450
https://inria.hal.science/inria-00078450

Submitted on 6 Jun 2006

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.

Browsing graphics without prior knowledge *

Daniel Zuwala
LORIA-INPL
BP 101
54602 Villers-lès-Nancy Cedex, France
Daniel.Zuwala@loria.fr

Jan Rendek
France Télécom R&D & LORIA
BP 101
54602 Villers-lès-Nancy Cedex, France
Jan.Rendek@loria.fr

Abstract

In this paper, we present a system to browse a set of graphical documents without prior knowledge of their content. This system relies on a new structural method locating regions potentially embedding a symbol, using an unsupervised reconstruction of the documents from their representation as chain of points. The method is complemented by a querying mechanism that retrieves from these regions those that are the most similar to a given input symbol. This mechanism uses a specific user feedback scheme to improve the accuracy of the retrieval.

1 Introduction

The recognition of graphical symbols has been subject to much effort throughout the years. The methods used include template matching, grammar-based matching, recognition techniques based on structural features or dynamic programming, and a number of structural methods based on graph matching.

One of the reasons for which symbol recognition is in many cases a very difficult and ill-defined problem is the large number and variety of symbols to be recognized. Except in strongly context-dependent applications, it may often be impossible to provide a base of all possible symbols.

The classical scenario generally consists in identifying a finite set of known symbols in a set of documents. The segmentation of the symbols is performed using *a priori* knowledge on the document [1], and the recognition method will integrate built-in knowledge of the possible symbols deformations, or will be trained on an extensive training set. Various efficient techniques have been developed, either using structural pattern description, or statistical pattern recognition techniques [2, 3, 7].

*This work is partially funded by a CIFRE contract between France Télécom R&D and INPL-LORIA.

In this article, we address a slightly different issue. We propose a method to locate and recognize symbols in a set of graphical documents with no *a priori* knowledge on the symbols. This method has been successfully implemented as a system allowing a user to browse through a set of technical drawings.

Our system is composed of two parts. The first is a structural method, based on the chain point representation of the documents. It generates hypothesis on the probable location of symbols. The second is a query system that selects among these hypothesis those that are most likely to represent a symbol similar to a candidate symbol inputted by the user.

The rest of the paper is organized as follows. In Section 2, we present the method used to detect the symbols embedded in the documents. The query system and the feedback scheme are detailed in Section 3. Section 4 describes the experiments used to test and to assess the performance of our system. Finally, conclusions and future work are given in Section 5

2 Symbol Extraction

Our method to locate symbols is based on the chains of points extracted from the documents. A chain of points includes connected points having only two neighbors, and its extremities are either junction points (more than two neighbors) or terminal points (one neighbor). The chains are extracted from the documents using several preprocessing steps including binarization, and skeletonization.

The goal of our method is to isolate sets of chains that may represent a symbol. A brute force approach would consider all possible ensembles of chains of a document as a potential symbol. This solution is, of course, inapplicable, first due to the resulting combinatorial explosion, but foremost because it would generate too many hypothesis that would hinder the performance of the retrieval.

Our idea is to merge iteratively these chains, in order to reconstruct the symbols belonging to the document. Once

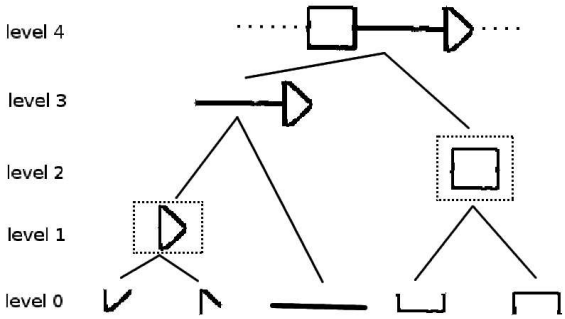


Figure 1. A Dendrogram representation

two chains have been merged, they are considered as a whole and cannot be separately be merged with others. The merging process from all the chains in the document can be represented by a dendrogram. Figure 1 exhibits such representation for a simple document composed of two basic symbols (a triangle and a square) connected by a line. The nodes enclosed by a dashed box are those corresponding to the reconstructed symbols.

What we need is to define an aggregation criterion, allowing us to decide which chains are to be merged at each step. Each step will represent a potential symbol. Well defining this criterion is critical, as we have to merge first the chains belonging to the symbols.

Defining the aggregation criterion. To define an efficient aggregation criterion, we made the two following assumptions on the representation of symbols.

- a symbol is a set of compact connected chains.
- a symbol tends to form a closed area, the chains composing it folding on each other.

From these two assumptions, we derived two measures for a chain of points, the compactness measure (noted m_c) and the folding measure (noted m_f), quantifying these criteria for a chain.

The compactness measure for a chain of points C is defined as follows. Let x_C^G be the barycenter of all the points of C, computed using equal weights, let n_C be the number of points contained in C, let x_i be the i-th point of C, and let $d(x, y)$ be the euclidean distance between a point x and a point y:

$$m_c(C) = \frac{\sum_{i=1}^{n_C} d(x_i, x_C^G)}{n_C} \quad (1)$$

To quantify the *folding* of a chain, we relate it to the area the chain encloses following the horizontal, or the vertical direction. An example of the computed area is given by Figure 2. The folding measure of the chain is defined as the



(a) the chain point (b) the area considered

Figure 2. The folding measure computation

ratio between this area and the bounding box enclosing the chain:

$$m_f(C) = \frac{Area_Consid ered(C)}{BoundingBox_Area(C)} \quad (2)$$

This gives us a measure ranging in $[0, 1]$.

These two criteria are then combined into a unique measure:

$$m(C) = m_c(C)(1 - m_f(C)) \quad (3)$$

For a chain C, the more compact the chain is, the lower $m_c(C)$ will be, and the more the chain encloses a region, the higher $m_f(C)$ will be. Thus, the lower $m(C)$ is, the more the chain C follows the assumption we made on a symbol representation. This measure will be used as the aggregation criterion in the construction of the dendrogram.

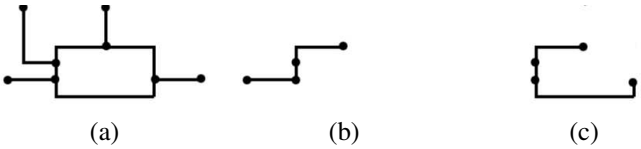
Look ahead improvements. In highly connected documents, building the dendrogram considering only the resulting aggregation measure of the resulting chain may yield to miss some highly connected symbols. This is due to the local view we have, which prevents us to choose a less relevant merge that would however lead to a better merge. To better this, we introduced a kind of look-ahead scheme. Instead of deciding which chains are to be merged, based only on the measure of the resulting chain, we use a modified criterion the consequences of that decision in the future merges that takes into account. In other words, we consider the chain itself, plus all the chains that are transitively connected to it, up to a fixed maximum depth. Let n_{max} be this maximum depth.

For each n in $[0..n_{max}]$, we compute m of each set of chain points. Let m_{min}^i be the minimum measure at depth i. We then combine these minimums into a single measure as follows.

$$m^{n_{max}}(C_1, C_2) = \sum_{k=0}^{n_{max}} \alpha(k) m_{min}^k(C_1, C_2) \quad (4)$$

where α is an attenuation factor.

Figure 3 shows the gain of this look-ahead method. (a) is the original symbol in the document. (b) is the merge we found from a local view. (c) is the merge we found with a more global view.



(a) (b) (c)

Figure 3. Lookahead demonstration

3 Document browsing

The process described above yields a set of potential symbols. Each node of the dendrogram is considered as a probable symbol. This decomposition is used as an input to a system that allows a user to retrieve all symbols similar to a given candidate symbol, or query. The query can either be drawn, or isolated from the documents.

With no *a priori* knowledge on the symbol, finding a descriptor and a metric to map the user concept of similarity into a distance measure between low level features is nearly impossible. Our idea is to use robust yet possibly sub-optimal descriptor and couple it to a relevance feedback scheme to better the exploration process.

Relevance feedback has been successfully applied in Content-Based Image Retrieval systems to improve the accuracy of the retrieval [10]. Its purpose is to involve the user in an interactive discussion loop, in order to adapt the similarity measure computed between two patterns to the user similarity concepts, based on their low-level representation. Many schemes have been devised to integrate the user feedback, either based on optimization techniques [8], or by casting the problem as a two-class classification issue and adapting known classification methods (see [11, 12] for comprehensive review). Here, we used a method based on the nearest neighbor classifier adapted from [4].

Preprocessing and features extraction. First all reconstructed symbols are filtered, using criteria based on the query area and aspect ratio. Then the Angular Radial Transform (ART) [6] is computed for all remaining samples. This descriptor is relatively fast to compute, and is robust to some degree to deformation, and distortion. However, our system can easily integrate any other descriptor.

Relevance feedback. From this initial query, the n best potential symbols retrieved, by computing the euclidean an distance from their representation to the query. A list of these symbols, ordered by their increasing distance to the query is presented to the user. Within this list, the user then selects which samples are relevant and which are irrelevant according to his own perception of similarity.

Then the relevance of each unlabeled samples is estimated from the minimal distance from its representation and the nearest relevant sample, and from the minimal distance from its representation and the nearest unrelevant sample. The closest it is to a relevant sample, the highest its own relevance is. On the contrary, the closest it is to a negative sample, the bigger the penalty assigned to its relevance is. The relevance R of a symbol s is computed as follows: let $\mathcal{N}$ be the set of negatively labeled symbols and $\mathcal{P}$ the set of positively labeled symbols:

$$\mathsf{R}(s) = 1 - e^{-\frac{d_{\min}(\mathcal{N},s)}{d_{\min}(\mathcal{P},s)}} \quad (5)$$

where $d_{\min}(\mathcal{X}, x)$ is the minimum distance from a set $\mathcal{X}$ to a symbol x.

The reconstructed symbols are sorted by decreasing relevance and submitted to the user, who can then again mark positive and negative samples, and loop over the feedback step, until he is satisfied with the result.

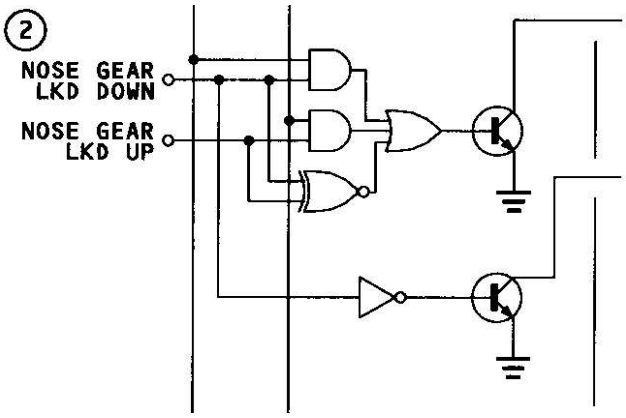


Figure 4. Test document example

4 Experiments

Experimental setup. We have tested our method with an image base of 30 documents of an airplane electric scheme. These documents are 3000x3000 pixels grayscale images. Figure 4 shows a part of such document. A document is first binarized, then a text/graphics separation is performed to remove all chains belonging to characters. Finally the skeleton image is computed using the (3,4)-Distance Transform [9], and the points chains are extracted.

From these document a groundtruth was built, consisting in 108 symbols belonging to 7 different classes.

We use *Precision* and *Recall* to evaluate the performance of the proposed approach. *Precision* is defined as the number of retrieved relevant samples over the total number or samples. *Recall* is defined as the number of retrieved relevant samples over the total number of relevant samples.

Each relevant symbol is used in its turn as initial query to the system. Two feedback iterations are performed, label-

Recall	25%	50%	75%	100%
0 rf	1.00	0.81	0.70	0.52
1 rf	1.00	0.96	0.86	0.70
2 rf	1.00	1.00	1.00	0.75

(a) Precision/Recall using ART

Recall	25%	50%	75%	100%
0 rf	0.91	0.60	0.31	0.15
1 rf	1.00	0.90	0.78	0.30
2 rf	1.00	1.00	1.00	0.40

(b) Precision/Recall using Zernike's moment

Figure 5. Average precision/Recall results

ing 10 symbols per iteration. To further test our feedback mechanism, we also used the Zernike moments [5] on the samples and used them as descriptor in place of the ART. The average precision for all queries at 25%, 50%, 75% of recall were computed and shown in Figure 5a for the ART, and Figure 5b for Zernike.

Results and observations. The decomposition of the 30 documents takes about 10 min (546s) on a Pentium4 2.6GHz. About 85,000 potential symbols were isolated, and only a third was kept after filtering.

The experiment reveal that:

- After filtering the hypothesis, a large number still remain as input for the query system. The 108 symbol instances generated about 27 000 possible symbols.
- The aggregation criterion we define was efficient for correctly reconstructing the symbols from the point chains. For the 108 symbol occurrences, only 4 had no matching reconstructed sample after the document decomposition.
- Though the ART is a well-suited descriptor for our experimental set (.81 precision for a 50% recall), the relevance feedback really boosted the precision of our system. The impact of the feedback is even more visible with a less efficient descriptor as Zernike, as precision was boosted from .31 to 1.0 at 50% recall.

5 Conclusion

In this paper, we introduced a new method to spot symbols in a set of graphical documents without any *a priori* knowledge. The only assumptions we made are that symbols form a compact set of connected chains, and that these chains tend to fold up. Coupled with a querying mechanism that use a statistical pattern descriptor and a feedback scheme, this method allows a user to browse a set of scanned document. The derived implementation yields excellent promising results, in terms of spotted symbols and precision of the result set for a query.

At present we plan on testing our system on bigger and more diverse document sets. We are working on defining smarter filtering rules to avoid the computational explosion due to the number of hypothesis generated. Better feedback schemes may also be required to minimize the user interaction, and to optimize the quality and the convergence of the recall.

References

[1] J.-P. Armand. Musical Score Recognition: A Hierarchical and Recursive Approach. In *Proceedings of 2nd International Conference on Document Analysis and Recognition, Tsukuba (Japan)*, pages 906–909, 1993.

[2] A. K. Chhabra. Graphic Symbol Recognition: An Overview. In K. Tombre and A. K. Chhabra, editors, *Graphics Recognition—Algorithms and Systems*, volume 1389 of *Lecture Notes in Computer Science*, pages 68–79. Springer-Verlag, Apr. 1998.

[3] L. P. Cordella and M. Vento. Symbol recognition in documents: a collection of techniques? *International Journal on Document Analysis and Recognition*, 3(2):73–88, Dec. 2000.

[4] G. Giacinto and F. Roli. Nearest-prototype relevance feedback for content based image retrieval. In *Proceedings of the 17th International Conference on Pattern Recognition, Cambridge (UK)*, pages 989–992, Aug. 2004.

[5] A. Khotanzad and Y. H. Hong. Invariant Image Recognition by Zernike Moments. *IEEE Transactions on PAMI*, 12(5):489–497, 1990.

[6] W.-Y. Kim and Y.-S. Kim. A new region-based shape descriptor. In *TR 15-01, Pisa (Italy)*, Dec. 1999.

[7] J. Lladós, E. Valveny, G. Sánchez, and E. Martí. Symbol Recognition: Current Advances and Perspectives. In D. Blostein and Y.-B. Kwon, editors, *Graphics Recognition – Algorithms and Applications*, volume 2390 of *Lecture Notes in Computer Science*, pages 104–127. Springer-Verlag, 2002.

[8] Y. Rui and T. Huang. Optimizing Learning in Image Retrieval. In *Computer Vision and Pattern Recognition*, page 1236, June 2000.

[9] G. Sanniti di Baja. Well-Shaped, Stable, and Reversible Skeletons from the (3,4)-Distance Transform. *Journal of Visual Communication and Image Representation*, 5(1):107–115, 1994.

[10] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on PAMI*, 22(12):1349–1380, Dec. 2000.

[11] H.-J. Zhang, Z. Chen, W.-Y. Liu, and M. Li. Relevance feedback in content-based image search. *World Wide Web*, 2:131–155, 2003.

[12] X. Zhou and T. S. Huang. Relevance feedback in image retrieval: a comprehensive review. *Multimedia Systems*, 8:536–544, 2003.