

Logo Spotting by a Bag-of-words Approach for Document Categorization

Marçal Rusiñol and Josep Lladós
Computer Vision Center, Dept. Ciències de la Computació
Edifici O, Universitat Autònoma de Barcelona
08193 Bellaterra (Barcelona), Spain
{marcal,josep}@cvc.uab.es

Abstract

In this paper we present a method for document categorization which processes incoming document images such as invoices or receipts. The categorization of these document images is done in terms of the presence of a certain graphical logo detected without segmentation. The graphical logos are described by a set of local features and the categorization of the documents is performed by the use of a bag-of-words model. Spatial coherence rules are added to reinforce the correct category hypothesis, aiming also to spot the logo inside the document image. Experiments which demonstrate the effectiveness of this system on a large set of real data are presented.

1. Introduction

Companies deal with large amounts of paper documents in daily workflows. Incoming mail is received and has to be forwarded to the correspondent addressee. A study on the invoice processing in several German companies [3] revealed that in average the cost of manually process (opening, sorting, internal delivery, data typing, archiving) these incoming documents is about 9€ per invoice. These costs represent an important quantity of money if we consider the amount of documents received by a big company at the end of the day. The Document Image Analysis and Recognition (DIAR) field has devoted, since its early years, many research efforts to deal with these kind of document images. In the first years, most of the contributions in this field only relied on the analysis of the text contained in the documents. Research was mainly centered in the analysis of textual documents and the design of automatic reader systems. Commercial OCR systems are currently working with very high recognition rates, and several systems to automatically process incoming documents have been designed. As an example, Viola et al. presented in [9] a system aiming to automatically enrout incoming faxes to the correspondent

recipient. However, all these systems only process typewritten information making the assumption that the recipient information is printed in the document image.

In many cases, graphic elements present in the documents convey a lot of important information. For instance, if a company receives a document containing the logo of a bank, usually this document should be forwarded to the accounting department, whereas if the document contains the logo of a computer supplier, it is quite probable that the document should be addressed to the IT department. The recognition of such graphic elements can help to introduce contextual information to overcome the semantic gap between the simple recognition of characters and the derived actions to perform brought by the document understanding. In this paper we use the presence of graphical elements (such as logos) to categorize the class of the incoming documents.

Within the DIAR field, the Graphics Recognition (GR) community has faced for many years the problem of symbol recognition yielding to good recognition results even with the presence of noise and other distortions. However, as pointed by Tombre and Lamiroy in [7], some challenges remain in this domain. The systems scalability is one of the main concerns. Usually, recognition schemes rely on a learning stage and then a classification strategy is used to recognize the input graphics. In that scenario, it is usual that the recognition ability of the system is severely impaired as the number of considered model classes grows. On the other hand, another open issue on this domain is the management of graphics that appear inside real documents. Many contributions in the literature, e.g. [10], that deal with logo recognition and retrieval, just focus on isolated or pre-segmented graphic images which are affected by synthetic noise and deformation sources. As noted in [8], one of the big challenges for the next years for the GR community is the localization/recognition of graphic symbols appearing in complete documents without any previous segmentation. To our best knowledge, in the literature, only Zhu and Doerman addressed in [11] the problem of logo spotting (i.e. the recognition and localization of logos in real documents) by

means of a cascade of classifiers. We present here a method which aims to categorize documents and to detect graphical logos in a single step. The main contribution of this paper is twofold. On the one hand the application itself and the use of well-known strategies of the Computer Vision field to this particular kind of images. On the other hand, the proposal of a segmentation-free recognition method which do not rely on a learning step but uses a single instance of logo models so as to benefit the scalability of the method.

The remainder of this paper is structured as follows: the next Section presents an overview of the proposed method. In Section 3, we detail the detection procedure from the feature extraction to the bag-of-words model used to categorize the documents. Section 4 focus on the addition of a set of spatial coherence rules which aim to refine the results and moreover, to perform logo spotting in addition to the categorization. Section 5 presents the experimental setup by using a large set of real documents. Finally, the conclusions and a short discussion can be found in Section 6.

2. Outline of the Approach

Our document categorization method is based on the presence of graphical logos in the incoming documents. This application can be seen as a particular case of the problem of object recognition of the Computer Vision field, where a certain object has to be identified within an image. However, the presented problem has certain particularities. First of all, the documents are in binary format and affected by the noise arising from the different acquisition systems. Local descriptors yield to poor feature vectors and the false alarms may increase. In addition, our application is intended to categorize an increasing number of document classes, and do not rely on a learning stage where several instances of the objects are shown. We propose a method that only requires one instance of the models to consider. Generally speaking, the presented method has a structure like the one proposed by Sivic et al. in [6], where a bag-of-words model is translated to the visual domain by the use of local descriptors over interest points.

We can see an overview of the presented method in Figure 1. The extracted local features from a document are matched against the codeword dictionary and an accumulator is used in order to decide to which category the queried document belongs. Let us further detail in the next Sections the followed steps.

3 Document Categorization by Logo Detection

The document categorization and the logo detection is performed by using a bag-of-words model of visual words.

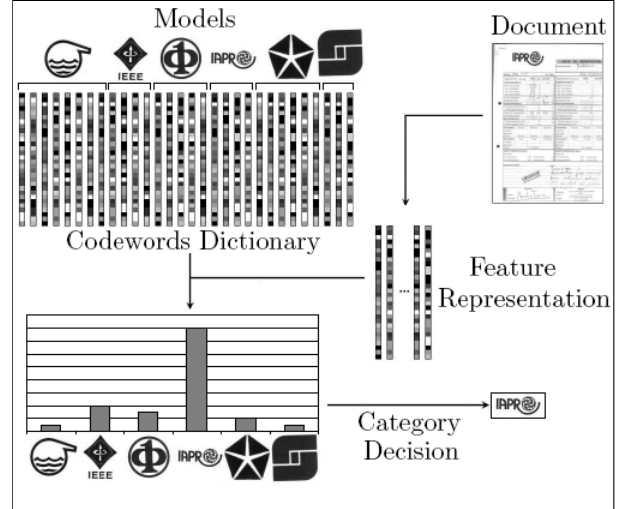


Figure 1. Method overview.

These visual words are described by local features. Let us first detail how these features are extracted and computed, and then focus on the bag-of-words model.

3.1 Feature Extraction and Description

Our method is inspired on the work presented in [1], focused on the recognition of trademarks in real images. In that work, the authors use SIFT features to match trademark models against video frames. We use a similar matching approach, whereas our aim is to categorize and to use several different logos as models. Logos are represented by a local descriptor applied to a set of previously extracted keypoints. These interest points are computed by using the Harris-Laplace detector presented in [5] which extracts points with high curvatures (as corners or junctions) and automatically selects the scale of the region to compute the local descriptor. A given logo L_i is then represented by its n_i feature points description:

$$L_i = \{(x_k, y_k, s_k, F_k)\}, \text{ for } k \in \{1 \dots n_i\}$$

where x_k and y_k are the x- and y-position, and s_k the scale of the k th key-point. F_k corresponds to the local description of the region represented by the key-point. An individual keypoint k of the logo L_i will be denoted as L_i^k . Our presented method is independent of the chosen descriptor. In our experiments, we used and compared the performance of two different local descriptors. On the one hand we use the SIFT features presented in [4] to describe the regions. On the other hand, we also used the shape context descriptor (SC) described in [2]. As we will see in the experimental results Section, each descriptor has its own strengths and weaknesses. The same notation applies when

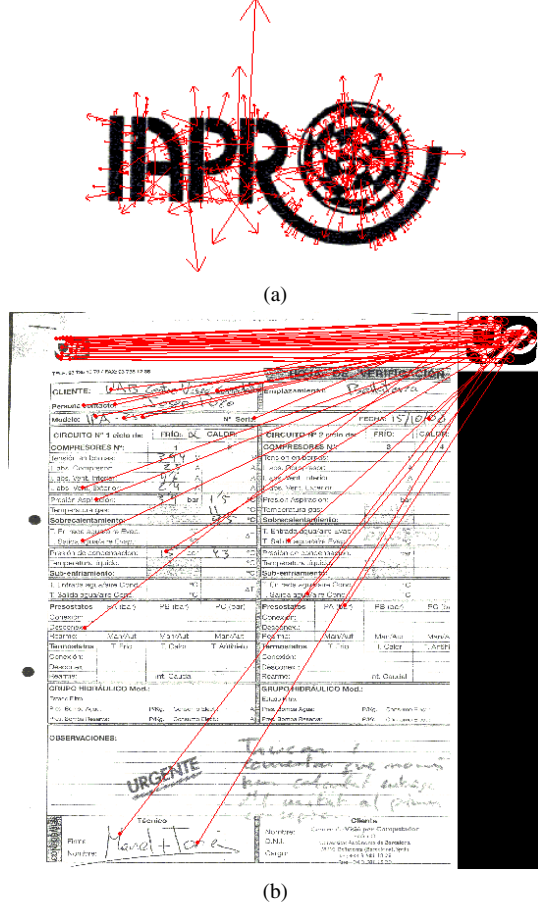


Figure 2. Matching logos in documents with the SIFT features. (a) SIFT features computed over an isolated logo model. (b) Matching between the model and the document.

the key-points and the feature vectors are computed over a complete document D_j . The matching between a keypoint from the complete document and the ones of the logo model is computed by using the two first nearest neighbors:

$$\begin{aligned} N_1(L_i, D_j^q) &= \min_k (F_q - F_k) \\ N_2(L_i, D_j^q) &= \min_{k \neq N_1(L_i, D_j^q)} (F_q - F_k) \end{aligned} \quad (1)$$

Then the matching score is determined as the ratio between these two neighbors:

$$M(L_i, D_j^q) = \frac{N_1(L_i, D_j^q)}{N_2(L_i, D_j^q)} \quad (2)$$

If the matching score M is lower than a certain threshold t this means that the keypoint is representative enough to be considered. By setting a quite conservative threshold

($t = 0.6$ in our experiments) we guarantee that the appearance of false positives is minimized since only really relevant matches are considered as so. We can appreciate in Figure 2 an example of the feature extraction and matching between a model and a document. However, for categorization purposes, we can not directly apply this matching procedure between the query document and all the model logos we consider. We use instead a bag-of-words model which have reached successful results for topic categorization. Let us describe in the next Section how we adapt this model to the visual domain.

3.2 Bag-of-visual-words

The Bag-of-visual-words is an analogy to the Computer Vision domain of the classic bag-of-words model, where a text is represented by an unordered set of words. In that case, an image is represented by collection of image patches. In our particular case, given a set of logo models considered as different categories, we extract all the feature vectors F_k^i from them. Each feature vector is associated to its corresponding logo model L_i . By joining all the feature vectors from all the logos, we obtain the codeword dictionary $W = [F_1^1, F_2^1 \dots F_k^i]$. This dictionary is computed off-line from all the model logo database. Given a query document D_j , all the feature vectors D_j^q are used as indexes and matched against the codewords of the dictionary W . The matching function M_q returns the index i corresponding to the logo class of the matched feature vector F_k^i as follows:

$$M_q = \{i | M(W, D_j^q) < t\} \quad (3)$$

Finally, the determination of whether a document contains a logo is done by using by accumulating hypothesis of document categories in an histogram H .

$$\begin{aligned} H[M_q] &= 0 \text{ at initialization,} \\ H[M_q] &= H[M_q] + 1, \text{ for } q \in \{1 \dots n_j\} \end{aligned} \quad (4)$$

The document category is finally determined by searching the maximum m of the accumulator H after normalizing each accumulation cell with the total number n_i of features of the corresponding logo k . If the value of m is less than a threshold T , which has been experimentally set, we consider that the document do not contain any logo and is categorized in a rejection class.

4 Spatial Coherence and Logo Spotting

Whereas bag-of-words models have been very successful in the text domain, the analogy to visual words for image categorization has an important drawback. Bag-of-words

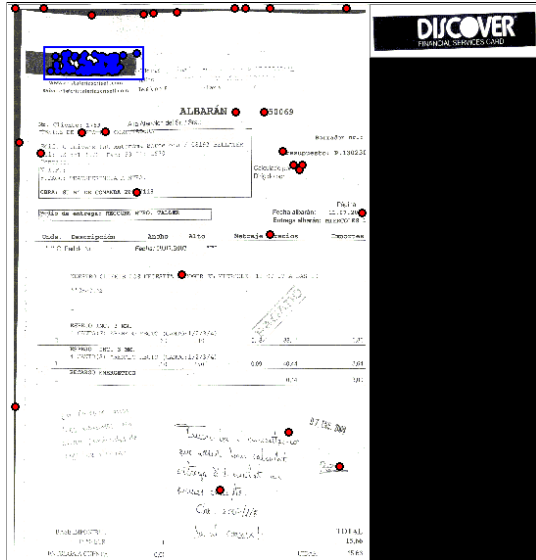


Figure 3. Introducing spatial coherence to spot logos.

models completely ignore the spatial relationship among features. Even if this drawback in the text domain is overcome due to the important impact of few keywords, in the image domain it is an important burden since the spatial layout among features has similar importance as the feature description itself.

To overcome this drawback we use a simple yet effective set of rules to guarantee that the spatial organization of features maintain certain coherence. Before contributing to the accumulator H , we get rid of the all the feature points of a same category i that are isolated in space by the use of the straightforward opening operation. By this means, we only consider clusters of keypoints which belong to the same category and which are close in space. As we can see in Figure 3, all the false alarms when matching keypoints are eliminated. The red dots are inconsistent hypothesis and the blue dots maintain a certain spatial coherence and are taken as likely hypothesis. The bounding-box of likely hypothesis are returned to the user as the zones of the document image where the logo should be found. The presented method, given a document is able to in a single step categorize it in a certain class and return the zone of the document which contain the logo.

5. Experiments

To provide a realistic evaluation of document categorization and logo spotting we used a large document collection. The collection consists in 1000 real document images which were sent by fax and then scanned. These images corre-

spond to several kinds of documents such as invoices, letters, receipts, etc. They contain both typewritten and handwritten text. Graphical elements such as logos, stamps, tables, etc. are also present in most of these documents. Typical dimensions of documents are near 2500×3500 pixels with varying resolutions and slight orientation changes. All the images were scanned in binary format by using the built-in thresholding method of the scanner. Ground-truth of the entire collection was manually created identifying 18 different logo classes appearing in nearly 180 images, the rest of document images do not contain any logo and are used to test if the presented method is also able to reject those documents.

5.1 Evaluation Methodology

The performance of categorization methods is usually evaluated by confusion matrices to see if the systems under evaluation confuse two classes, mislabelling one as another. In addition, the true positive rate (TPR) and false positive rate (FPR) are used as evaluation measures in order to compare the performance among different methods. These ratios are derived from the contingency table and defined in terms of the amount of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN):

$$TPR = \frac{TP}{(TP + FN)} ; FPR = \frac{FP}{(FP + TN)} \quad (5)$$

The TPR ratio measures the effectiveness of the system in retrieving the relevant items. Whereas the FPR ratio measures the probability that a non-relevant document is retrieved by the query. In our experiments we use the TPR ratio to summarize the correct categorization of documents containing a given logo. The FPR is used to measure the amount of documents that do not contain any logo which are incorrectly identified as belonging to a certain class.

5.2 Performance Comparison

Table 1. Evaluation Measures

Descriptor	TPR (%)	FPR (%)	Time (secs.)
SIFT	92.2	1	3.25
SC	81.6	0.3	1.34

We can appreciate in Figure 4 the obtained confusion matrices after running the whole experimental categorization. We can appreciate some differences between the use of SIFT features and SC. For example, when using SC, a lot of documents are incorrectly classified as class 8 (shown in

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	100			2,44														7,69	
2		100																	
3			100																
4				92,69				14,29										3,85	
5					90														
6						90,9													
7							71,42				50								
8								100											
9				4,87					80										
10					10					100									
11											50								
12												100							
13													100						
14														90					
15						9,1			20						100			3,85	
16																20		100	
17																			84,61
18																			100

(a) SIFT

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
1	100			2,44															3,84	
2		100																		
3			100																	
4				95,36										16,67		10		3,84		
5					60															
6						100								40				3,84		
7							57,15													
8								7,32	10	26,67	100	20					7,14	10	7,7	
9									4,88											
10										10										
11											0									
12												50	100							
13														66,66						
14															60					
15																65,72				
16																	80		3,84	
17																			69,26	
18																			3,84	88,88

(b) Shape Context

Figure 4. Confusion Matrices.

row 8), or the documents corresponding to class 17 are usually misclassified in other document categories (shown in column 17). Those classification errors do not correspond with similar logos designs. The misclassifications lead the overall TPR shown in Table 1 to be lower when using SC than when using SIFT. On the other hand, when we test the documents that do not contain any logo and should be categorized in the rejection class, the SIFT features perform worst than SC, as shows the FPR . In addition, the computational complexity when using SIFT is higher due to the highest number of dimensions of the feature vectors than when using SC, resulting in a higher querying time.

6 Conclusions

In this paper we have presented a method for document categorization in terms of the presence of a certain graphical logo. The use of a bag-of-words model reformulated to manage local features combined by a set of spatial coherence rules aim to spot the logo inside the document image in addition to determine the category of the queried document. The presented experiments demonstrate the effectiveness of the method on a large set of real document images. One of the most important current challenges of the GR community is the proposal of segmentation-free recognition methods that can analyze complete documents, logo spotting techniques should be further investigated.

Acknowledgments

This work has been partially supported by the Spanish projects TIN2006-15694-C02-02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

References

- [1] A. Bagdanov, L. Ballan, M. Bertini, and A. D. Bimbo. Trademark Matching and Retrieval in Sports Video Data-

- bases. In *Proceedings of the International Workshop on Multimedia Information Retrieval*, pages 79–86, 2007.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. *IEEE Trans. on Pattern Analysis and Machine Intel.*, 24(4):509–522, 2002.
- [3] B. Klein, S. Agne, and A. Dengel. Results of a Study on Invoice-Reading Systems in Germany. In *Document Analysis Systems VI*, volume 3163 of *Lecture Notes on Computer Science*, pages 451–462. 2004.
- [4] D. Lowe. Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [5] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
- [6] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering Objects and their Location in Images. In *Proceedings of the International Conference on Computer Vision, ICCV05*, pages 370–377, 2005.
- [7] K. Tombre and B. Lamiroy. Graphics Recognition - from Re-engineering to Retrieval. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition, ICDAR03*, pages 148–155, 2003.
- [8] E. Valveny, P. Dosch, A. Fornés, and S. Escalera. Report on the third contest on symbol recognition. In *Graphics Recognition. Recent Advances and New Opportunities*, volume 5046 of *Lecture Notes on Computer Science*, pages 321–328. 2008.
- [9] P. Viola, J. Rinker, and M. Law. Automatic Fax Routing. In *Document Analysis Systems VI*, volume 3163 of *Lecture Notes on Computer Science*, pages 484–495. 2004.
- [10] C. Wei, Y. Li, W. Chau, and C. Li. Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure. *Pattern Recognition*, 42(3):386–394, 2009.
- [11] G. Zhu and D. Doerman. Automatic Document Logo Detection. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition, ICDAR07*, pages 864–868, 2007.