Classification of Administrative Document Images by Logo Identification

Marçal Rusiñol[1], Vincent Poulain D'Andecy[2], Dimosthenis Karatzas[1], and Josep Lladós[1]

[1] Computer Vision Center, Dept. Ciències de la Computació
Edifici O, UAB, 08193 Bellaterra, Spain
{marcal,dimos,josep}@cvc.uab.cat
[2] ITESOFT
Parc d'Andron, Le Séquoia
30470 Aimargues, France
vincent.poulaindandecy@itesoft.com

Abstract. This paper is focused on the categorization of administrative document images (such as invoices) based on the recognition of the supplier's graphical logo. Two different methods are proposed, the first one uses a bag-of-visual-words model whereas the second one tries to locate logo images described by the blurred shape model descriptor within documents by a sliding-window technique. Preliminar results are reported with a dataset of real administrative documents.

Keywords: Administrative Document Classification, Logo Recognition, Logo Spotting.

1 Introduction

Companies deal with large amount of paper documents in daily workflows. Incoming mail is received and has to be forwarded to the correspondent addressee. The cost of manually processing (opening, sorting, internal delivery, data typing, archiving) incoming documents represents an important quantity of money if we consider the daily amount of documents received by large companies.

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. As an example, Viola and collaborators presented in [6] a system aiming to automatically route incoming faxes to the correspondent recipient. However, most of the systems only process typewritten information making the assumption that the provider information is printed and well recognized by the OCR engine.

In many cases, some graphic elements that are 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 logo images in order to categorize the incoming document as belonging to a certain supplier.

ADAO (Administrative Document Automate Optimization) is an FP7 Marie-Curie Industry-Academia Partnerships and Pathways (IAPP) project between the French company ITESOFT and the Computer Vision Center (UAB) in Spain, which is focused on key document analysis techniques involved in a document workflow management. Within this project, one of the tasks is centered on the categorization of document images based on trademark identification. In this paper we report the obtained results for this specific task. Two different methods have been proposed, the first one uses a bag-of-visual-words (BOVW) model whereas the second one tries to locate logo images described by the blurred shape model (BSM) descriptor within documents by a sliding-window technique.

The remainder of this paper is organized as follows: We detail in Section 2 the bag-of-visual-words model and in Section 3 the sliding window approach. Section 4 presents the experimental setup. Finally, the conclusions and a short discussion can be found in Section 5.

2 Bag-of-visual-words Classification

This first method is based on the work we presented in [5]. In the proposed approach, the bag-of-words model is translated to the visual domain by the use of local descriptors over interest points. Documents are thus categorized based on the presence of visual features coming from a certain graphical logo. We can see an overview of the presented method in Figure 1.

Logos are represented by a local descriptor applied to a set of previously extracted keypoints. The interest points are computed by using the Harris-Laplace detector presented in [4]. A given logo L_i is then represented by its n_i feature points description:

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

where x_k and y_k are the x- and y-position, and s_k the scale of the kth keypoint. F_k corresponds to the local description of the region represented by the key-point. In this case, we use the SIFT local descriptors presented in [3]. The same notation applies when 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:

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

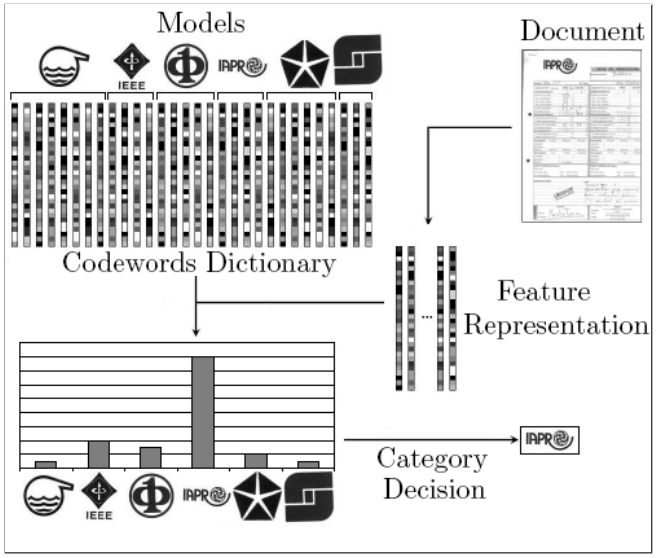


Fig. 1: Bag-of-visual-words model overview

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

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

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 such. That is, two keypoint descriptors are matched only if the ratio between the first and the second nearest neighbor is below a certain threshold. When a word in the dictionary belonging to a class C and a feature vector from the document are matched we accumulate a vote for the documents category C. After all the features of the document are processed, the class accumulating more evidences is the one selected as the document class.

3 Sliding Window over BSM Descriptors

The second method uses a sliding window framework together with the blurred shape model (BSM) descriptor [1] to categorize the incoming documents and locate the position of the logo.

The BSM descriptor spatially encodes the probability of appearance of the shape pixels and their context information in the following way: The image is divided in a grid of $n \times n$ equal-sized subregions, and each bin receives votes

from the pixels that fall inside it and also from the pixels falling in the neighboring bins. Thus, each pixel contributes to a density measure of its bin and its neighboring ones. The output descriptor is a histogram where each position corresponds to the amount of pixels in the context of the sub-region. The resulting histogram is L1-normalized.

In the original formulation of the BSM descriptor, pixel density was computed over a regular $n \times n$ grid, provoking that the shapes to compare have to be previously segmented. In our case we reformulate the BSM descriptor by forcing the spatial bins to have a fixed size (100x100 pixels in our experimental setup). Images of different size will result in feature vectors of different lengths. In order to locate a logo within a document image we use a sliding-window approach computed as a normalized two-dimensional cross-correlation (described in [2]) between the BSM description of the model logo and the BSM description of the complete document. By using this reformulation of the BSM descriptor, the chosen size of the buckets will define the level of blurring and subsequently the information reduction for both the logos and the documents.

As the result of the cross correlation between the BSM models and the BSM descriptor from the document, a peak should be formed in the location where there is a high probability to find a something similar to the given logo. This process is repeated for each logo in the knowledge database, and the peak having the highest response would be the best match between a certain zone of the document and the logo model, thus representing the most plausible class C of the document.

In order to increase the robustness of the method, we want to give the same importance to match "black" pixels and to match "white" pixels. To do so, the normalized cross correlation is computed for both the BSM description and the inverse of the BSM descriptions. In the final step, the probability maps coming from both normalized cross correlations are combined by multiplying them to get rid of the background noise.

One of the advantages of this method is that from the obtained probability maps, we can have not only the class of the document but also the location within the document where the most feasible logo is found. We can see an example of the whole procedure in Figure 2.

4 Experimental Results

4.1 Dataset

The selected dataset consists of 3337 TIF binary images of scanned invoices. From this collection, 204 different document classes identifying the invoice supplier have been determined. The ground-truthing protocol was the following. We first define with an annotation tool as many bounding boxes as logos in the document.

Segmenting a logo is somehow subjective and there are many cases where it is difficult to determine what a logo is. We followed these rules to produce the groundtruth:

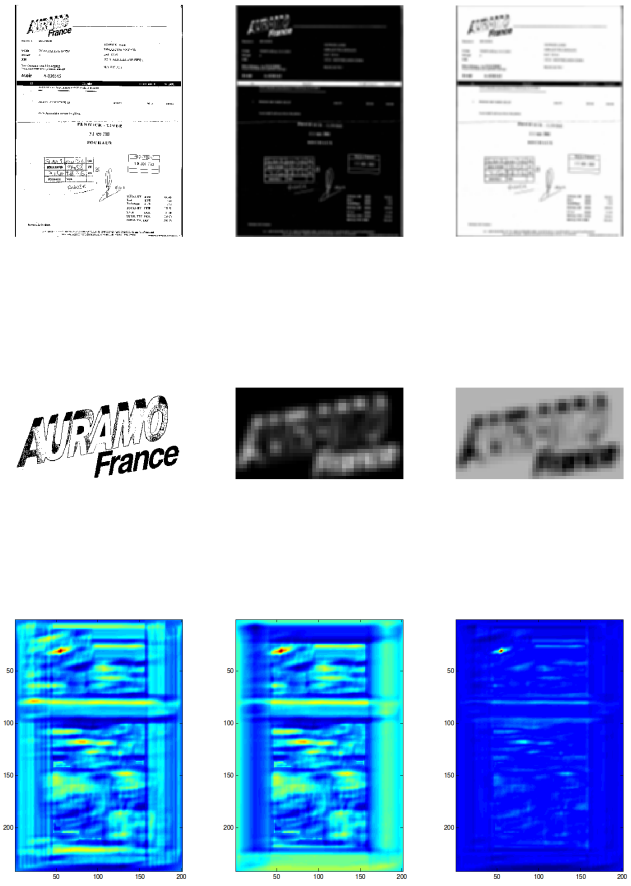
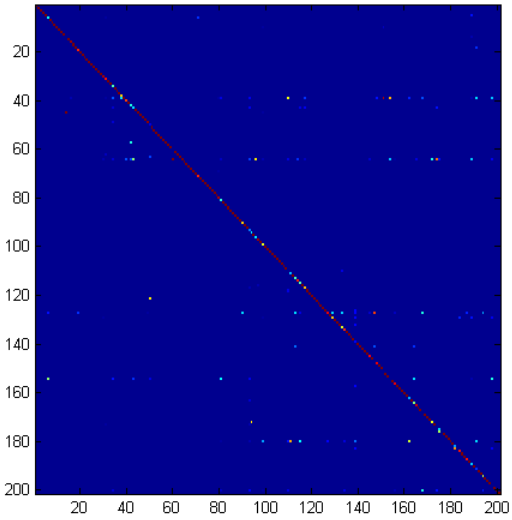
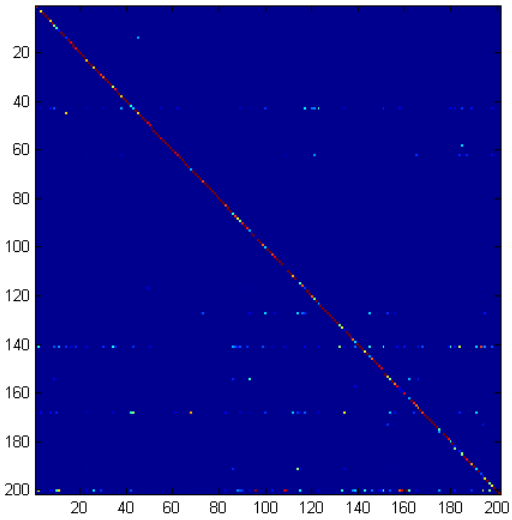


Fig. 2: Original images and BSM descriptors of the documents (first row) and logo models (second row). Probability maps for the BSM, the inverse BSM and the final combination of both are given in the third row.

- If there is some text close to the logo (usually the address), we tried not to select this text as a part of the logo.
- In some documents multiple logos might appear, we define a bounding box for each of them.
- If in the document we find multiple logos which are close to each other but are clearly of different nature, we try to define a separate bounding box for each of them.
- For the documents that do not contain any kind of graphical logos, we select the address as the logo of the document (see Figure 4). We keep track of these particular documents that do not contain any graphical logo.



a)



b)

Fig. 3: Confusion matrices for the a) BOVW and b) BSM methods when using 200 models.

Finally, the annotation tool returns an XML like file with the same name as the image file defining the location of the bounding box and the label for each of the bounding boxes.

4.2 Results

We present in Table 1 the results of the document classification for the two presented methods when considering a different amount of model logos. In this experiment, only the subset of the 3337 document images that correspond to these particular logos is used.

Table 1: Document classification

Dataset	BOVW	BSM
50 models / 902 documents	88.11	92.84
100 models / 1832 documents	90.45	89.79
200 models / 3295 documents	87.07	78.36

During the analysis of our results we realized that there were some logo designs that introduced much more noise when using them as a cue to categorize documents than others. These logo designs were the responsible for obtaining better performances in the BOVW scenario when considering 100 models that when considering 50 models. We can see this effect in the confusion matrices presented in Figure 3. Looking in detail at those logo designs we realized that most of the classes where we obtained poor performances corresponded at mostly-textual logos. We can see an example of these logo designs that we have in our dataset in Figure 4.

It is obvious that trying to recognize this kind of logo designs from a graphical point of view does not make much sense. We run an additional experiment with a reduced model dataset where we just included logos having a graphic-rich design. Some examples of these graphic-rich logo designs can be seen in Figure 5. The obtained results with these models are shown in Table 2.

Table 2: Document classification with only graphical logos

Dataset	BOVW	BSM
50 graphical models	87.86	99.55

We can see that the BSM method outperforms the BOVW method in this case. The BSM method is also much cheaper to compute than the BOVW. However, how these methods would scale when considering a larger amount of model logos is still an unanswered question that needs to be further investigated.

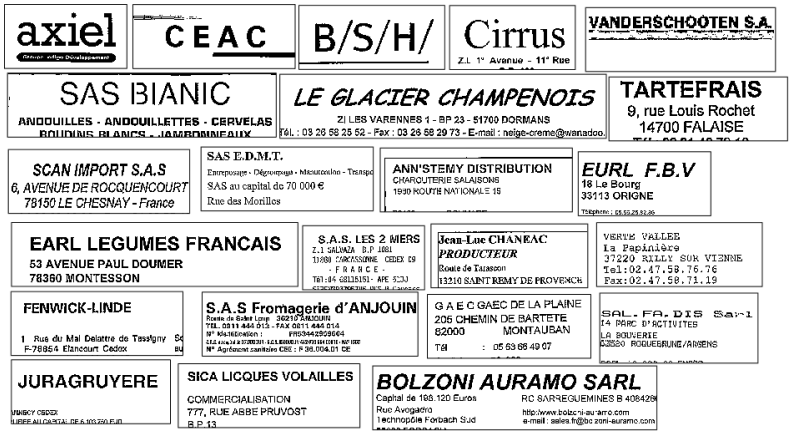


Fig. 4: Example of mostly-textual logos we have in the dataset.

Fig. 5: Example of graphic-rich logos we have in the dataset.

Another important issue is that with the BOVW model, all the spatial information is lost and we just obtain the category of the document as output of the system, whereas with the proposed approach based on cross-correlations over the BSM descriptors, not only we obtain the class of the input document but also the position of the logo in the document. In Figure 6 we show a screenshot of our classification demo software where we can see for an incoming document image, its recognized logo with its corresponding location in the original image.

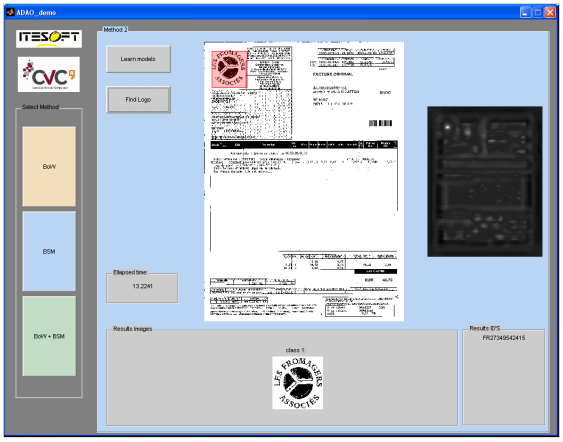


Fig. 6: Example of the logo localization when using the BSM descriptor.

5 Conclusion

In this paper we have presented and compared a couple of methodologies aiming to perform document classification in terms of the presence of a given logo image. The obtained results are encouraging even if they are reported in a low-scale scenario. It has been shown that to take into account graphical information can be very useful for document classification, at least for disambiguation in the cases where the answer of the main administrative document classifier has low confidence.

Acknowledgment

This work has been supported by the European 7th framework project FP7-PEOPLE-2008-IAPP: 230653 ADAO. The work has been partially supported as well by the Spanish Ministry of Education and Science under projects RYC-2009-05031, TIN2011-24631, TIN2009-14633-C03-03, Consolider Ingenio 2010: MIPRCV (CSD200700018) and the grant 2009-SGR-1434 of the Generalitat de Catalunya.

References

1. S. Escalera, A. Fornés, O. Pujol, A. Escudero, and P. Radeva. Circular blurred shape model for symbol spotting in documents. In *Proceedings of the IEEE International Conference on Image Processing*, pages 2005–2008, 2009.
2. J.P. Lewis. Fast normalized cross-correlation. In *Vision Interface*, volume 10, pages 120–123, 1995.
3. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

4. K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
5. M. Rusiñol and J. Lladós. Logo spotting by a bag-of-words approach for document categorization. In *Proceedings of the Tenth International Conference on Document Analysis and Recognition*, pages 111–115, 2009.
6. 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.