Improving Logo Spotting and Matching for Document Categorization by a Post-Filter based on Homography

Viet Phuong LE(1), Muriel VISANI(1), Cao De TRAN(2), Jean-Marc OGIER(1)
(1)Laboratory L3I, Faculty of Science and Technology, La Rochelle University, France
(2)College of Information and Communication Technology, Can Tho University, Vietnam
{viet_phuong.le, muriel.visani, jean-marc.ogier}@univ-lr.fr, tcde@cit.ctu.edu.vn

***Abstract*—Digital document categorization based on logo spotting and recognition has raised a great interest in the research community because logos in documents are sources of information for categorizing documents with low costs. In this paper, we present an approach to improve the result of our method for logo spotting and recognition based on keypoint matching and presented in our previous paper [7]. First, the keypoints from both the query document images and a given set of logos (logo gallery) are extracted and described by SIFT, and are matched in the SIFT feature space. Secondly, logo segmentation is performed using spatial density-based clustering. The contribution of this paper is to add a third step where homography is used to filter the matched keypoints as a post-processing. And finally, in the decision stage, logo classification is performed by using an accumulating histogram. Our approach is tested using a well-known benchmark database of real world documents containing logos, and achieves good performances compared to state-of-the-art approaches.**

***Keywords*—logo spotting, homography, pattern recognition, document analysis.**

I. INTRODUCTION

Logos (as well as seals) are very useful for the categorization of documents, especially in business and administrative documents. They allow us to quickly determine the source of the documents and accurately with low costs. In recent years, the explosion of the amount of digital documents poses challenges for the categorization and indexing of digital documents based on their origins.

As a consequence, many research works in the field of logo recognition have been carried out. In particular, Doermann et al. [5] use a combination of text, shape, and global and local affine invariants for logo recognition. Meanwhile, Zhu *et al.* [8] present an approach using a multi-scale boosting strategy to detect and extract logo(s) in document images. At a coarse image scale, a Fisher classifier provides an initial classification. Then, each logo candidate region is further classified at a finer image scales by a cascade of simple classifiers.

Jain and Doermann [13] present an approach for logo retrieval without segmentation. SURF features are used for logo retrieval; and they propose an indexing technique to group feature vectors and a filter method based on the properties of the features orientation and their geometric characteristics. Meanwhile, Rusinol and Llados [14] introduce a method for organizing and indexing logos based on describing logos by a variant of the shape context descriptor.

In another paper of Rusinol and Llados [6], they propose a logo spotting method where the logo image and the query documents image are described by a set of SIFT features describing keypoints. A bag-of-word model is further used for matching. In order to filter the matching keypoints and consider only the keypoints belonging to the logo in the query document, they consider clusters of keypoints.

In our previous work [7], we present an approach for logo spotting and recognition for documents categorization based on keypoint matching and density-based clustering as a post-processing for an accurate localization of the logo. First, similarly to the work by Rusinol and Llados [6], the keypoints from both the query document images and a given set of logos (logo gallery) are extracted and described by Scale Invariant Feature Transform (SIFT) descriptor [3]; then, keypoints are matched in the SIFT feature space using the nearest neighbor rule, with ambiguity rejection based on the two nearest neighbours. Second, the matched keypoints are clustered by a density-based clustering algorithm. In our approach, we use the Density-Based Spatial Clustering of Application with Noise (DBSCAN) [9]. We consider the cluster containing the maximum number of matched keypoints as a logo region candidate. Finally, in the decision step, we use a normalized accumulating histogram to compute a dissimilarity measure between the logo region candidate and each logo in the gallery and provide the final decision. This approach produces certain good results when compared to other approaches (Table 2). However, not all the matched pairs of keypoints in each cluster are correct because our previous method does not integrate enough information concerning the spatial distribution of the matched keypoints. Incorrect matches (inconsistent with the spatial distribution of the keypoints in the logo) need to be rejected before the final decision step.

The contribution of this paper is to add to our previous approach an additional third step in which the pairs of incorrectly matched keypoints are filtered based on homography using RANSAC (see Figure 1).

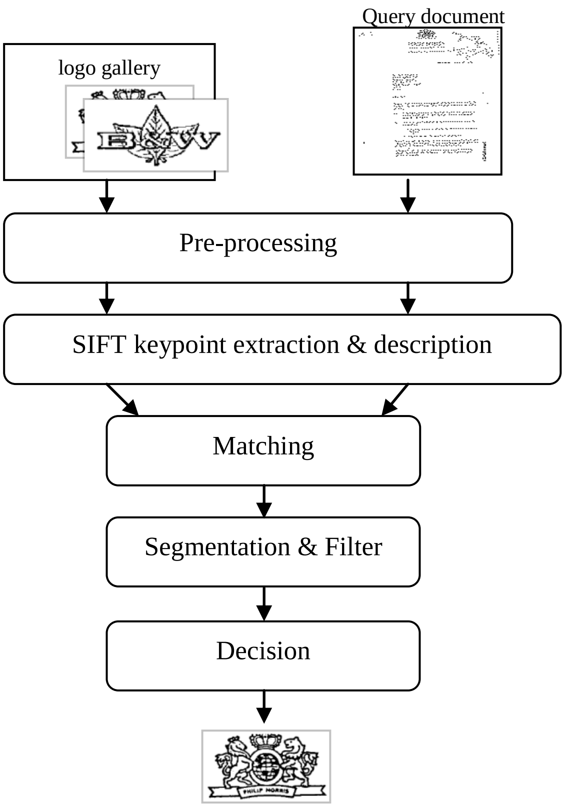


Figure 1: The outline of our approach

This paper is organized as follows. In section II., we describe how to estimate the homography using RANSAC method. In section III., we report our method to filter the matched keypoints using the homography. Finally, we present the experimental results in section IV. and draw the conclusions in the last section.

II. HOMOGRAPHY USING RANSAC

An homography is an invertible transformation from points in R^2 to points in R^2 that maps lines to lines [12]. Homography is used in many applications relying on geometry. In the field of computer vision, homography can be used for computing matches between images [12]. In a 2D plane, let us consider a set *S* of source points in the original plane $s_i(x'_i,y'_i)$, and a set *T* of target points $t_i(x_i,y_i)$ in the target plane. Homography is a perspective transformation *H* between the source and the target planes:

$$s_i \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \sim H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

where *H* is the 3x3 homography matrix:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

and the back-projection error (for projecting the target points in the original plane) is:

$$\sum_i \left(x_i' - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2 + \left(y_i' - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2$$

We need to estimate the homography H so that the back-projection error is minimized. There are many algorithms to estimate the homography H. The RANdom SAmple Consensus (RANSAC) algorithm, presented by Fischler and Bolles in [4], is a very robust algorithm for estimating H. It can deal with a large number of outliers. The main idea is to identify the outliers as data samples with greatest residuals with respect to the fitted model. The steps of the general RANSAC algorithm are as follows [12]:

1. Randomly select a subset *s* of *n* points of *S* and estimate the homography model from *s* and the corresponding subset *t* of matched keypoints in *T*.

2. Determine the set of inliers and outliers based on a distance threshold θ to the model.

3. If the number of inliers is greater than some threshold Θ, re-estimate the model and terminate.

4. If the number of inliers is less than Θ, then select a new subset *s* and repeat.

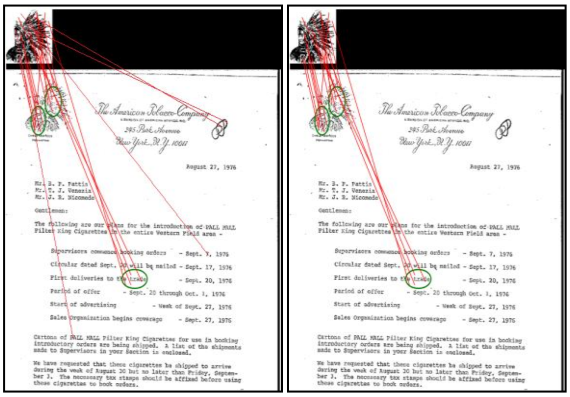
5. After repeating N times, the largest consensus set of inliers is selected, and the model is re-estimated using this set.

The parameters of this algorithm are discussed in [12]. We set the values of these parameters using experiments.

In our application, we try to compute a homography matrix H from the matched keypoints in the logo gallery to its corresponding logo in the query document. Therefore, each logo in the gallery is considered as a source plane while the target plane consists in the candidate logo region (biggest cluster obtained by applying DBSCAN).

III. Filter by Homography

After the steps of pre-processing and SIFT keypoint extraction, description and matching (see Figure 1), not all the matched keypoints are located inside the logo region. Some of them are located in other regions of the document image. We consider these keypoints as noise, and we do not want them to be taken into account in our final decision.



(a) (b)

Figure 2. Before (a) and after (b) grouping and segmentation by DBSCAN

Most of those incorrectly matched keypoints are rejected after clustering by DBSCAN (see Figure 2). However, a few pairs of keypoints which are incorrectly matched (see Figure 3a, b, and c) remain. We propose an algorithm to solve this problem based on homography using RANSAC. We try to find a transformation between the pairs of matched keypoints in each cluster to integrate spatial relationships between the keypoints in the logo and in the clusters. And then compute the transformation again after determining all matched keypoints in a bounding box covering the logo region candidate. This last step aims at solving cases where a cluster region may not cover the logo region, but just a part of the logo region (for an example see Figure 2b, where there are two clusters in the logo region).

Our algorithm is as follows:

For each logo region candidate and each logo in the gallery:

1. Find a transformation H between the pairs of the matched keypoints in the logo region candidate. Let $s_i(x_{i1},x_{i2})$ be the coordinates of the matched keypoints in the logo gallery, and let $t_i(y_{i1},y_{i2})$ be the coordinates of the matched keypoints in the query document image (Figure 3a).

2. Determine a bounding box which may contain a logo in the query document, thanks to the transformation H and the four corners of the minimal bounding box of the logo in the logo gallery (Figure 3b).

3. Re-estimate the transformation H using all the pairs of matched keypoints in the bounding box (Figure 3c).

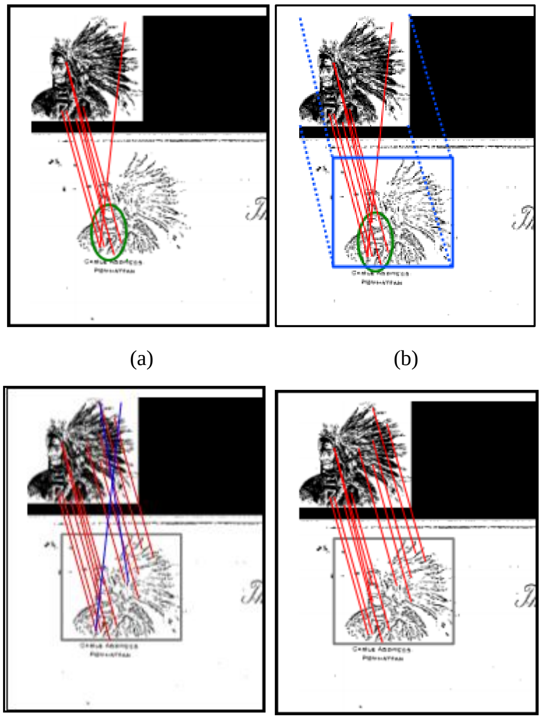
4. Filter the incorrectly matched keypoints: if $\| t_i - H(s_i) \| >= \theta$ then reject this pair (Figure 3c & 3d).

Finally, like in [7], we use a normalized accumulating histogram for the final decision. First, we count the number of matched keypoints in the cluster containing the highest number of matched keypoints and put the number corresponding to logo i into the i^{th} cell of the accumulating histogram. Then, we normalize each cell with the number of keypoints in the corresponding logo. The final decision is made by searching the maximum m of the accumulating histogram thanks to a threshold T (Figure 4): if m is equal or greater than the threshold T, it means that the document image contains this logo. Otherwise, it does not contain any logo from the gallery (or no logo at all).

IV. Experiments

Like in our previous work, we use the Tobacco-800 dataset [1]. We perform two series of experiments. The first one aims at evaluating the performances of our algorithm for logo recognition.

The second experiment aims at comparing our approach with other state-of-the-art methods and our previous approach for logos detection. According to [9], estimating automatically the values of parameters ε and *MinPts* of DBSCAN is very difficult. The values of the DBSCAN parameters ε and *MinPts* are identical to our previous approach, ε=60 and *MinPts*=5. For homography using RANSAC method, according to [12], we choose the threshold θ=6 and the other parameters are estimated based on the number of data points (using default parameters).



(a) (b) (c) (d)

Figure 3. The filter with transformation in our approach. Blue lines are the pairs of matched keypoints which are rejected.

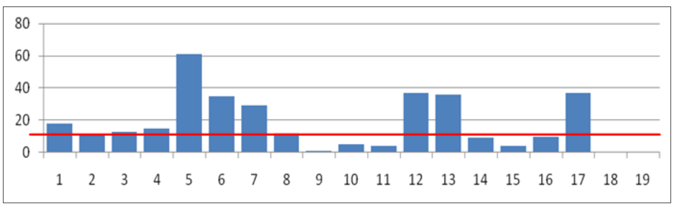


Figure 4. The accumulating histogram. Red lines show the threshold T. In this case, our algorithm decides that the document contains logo number 5

To evaluate our results, we use three evaluation metrics: *accuracy*, *precision* and *recall*. In our first experiment, *accuracy* is the ratio of correct results amongst the 374 document images containing such logos. In our second experiment, if we denote the class “logo” as positive and the class “no logo” as negative, *accuracy* is the proportion of correct results (True Positives + True Negatives) over all the test document images. *Precision* is the proportion of True Positives over all the positive results while *recall* is the proportion of the True Positives over all the document images containing a logo (positives).

Table 1. Performance evaluation and comparison for logo recognition.

ID	Approaches	Accuracy
01	V.-P. Le *et al.* [7]	86.90%
02	Our approach without DBSCAN	55.61%
03	Our approach with DBSCAN	88.77%

In the first experiment, we consider the 15 logos from the database contained in at least 3 document images, all the 374 document images containing such logos and all the 878 document images in the database without any logo. We compare the recognition rates we obtain with our previous approach in [7] and our current approach but without the DBSCAN segmentation. The results shows an improvement, as the accuracy reaches 88.77%, instead of 86.90% (see Table 1).

Table 2. Performance evaluation and comparison for logo detection.

ID	Approaches	Accuracy	Precision	Recall
01	G. Zhu and D. Doerman [8]	84.2%	73.5%	-
02	Z. Li *et al.* [10]	86.5%	99.4%	-
03	T.-A. Pham *et al.* [11]	91%	85%	-
04	V.-P. Le *et al.* [7]	94.04%	91.11%	88.94%
05	Our approach without DBSCAN	85.61%	90.61%	58.42%
06	Our approach with DBSCAN	95.86%	97.67%	88.42%

In the second experiment dedicated to logo detection, we compare the precision and accuracy measures of our approach with our previous approach and other state-of-the-art methods. Please note that this comparison has to be considered carefully, as in our approach and in our last approach we know the logo gallery, which is not generally the case for pure logo detection approaches. Table 2. shows an increase in Accuracy and Precision thanks to the filter by homography using RANSAC. However, the recall drops slightly of 0.52% because this filter also reduces the number of the matched keypoints in each cluster for the decision step; as a result, a few document images containing logo are incorrectly classified as containing no logo.

V. CONCLUSION

In this paper, we present an approach to improve logo spotting for document categorization based on post-filtering by homography using RANSAC. We use the SIFT features of keypoints for describing the query document images and the logos in the gallery. Keypoint matching is determined by using the nearest neighbor rule, with ambiguity rejection based on the two nearest neighbours. To locate candidate logo regions in the query document, we apply segmentation using DBSCAN density-based clustering. The originality in this paper compared to our previous work is that we further use homography using RANSAC to filter the incorrectly matched keypoints both at the cluster level and at a coarser level that might include multiple clusters. The addition of this intermediate step before final decision (based on normalized accumulating histograms) enhances our recognition results. Finally, the document categorization is determined by labeling the query document with the logo in the gallery corresponding to the maximum value in the normalized histogram. The current version of our algorithm cannot handle more than one logo, as we consider only the biggest logo for RANSAC filtering. But, we are currently working on its extension dealing with multiple logos, by simply considering every cluster in the document image and applying the same procedure.

We are also investigating document retrieval, where the logo is the query and the gallery contains the documents. We also study different strategies to better integrate spatial knowledge in the keypoint matching algorithm so as to enhance the precision of our system and its performances in a multi-scale context.

REFERENCES

[1] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J. Heard.*Building a test collection for complex document information processing*. In Proc. 29th Annual Int. ACM SIGIR Conference (SIGIR 2006), pp. 665-666, 2006.

[2] G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger. *Multi-scale Structural Saliency for Signature Detection*.In Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2007), pp. 1-8, 2007.

[3] D. Lowe. *Distinctive Image Features from Scale-invariant Keypoints*. International Journal of Computer Vision, 60(2): pp. 91-110, 2004.

[4] M. A. Fischler and R. C. Bolles. *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. Graphics and Image Processing, vol. 24: pp 381-395, 1981.

[5] D. Doermann, E. Rivlin& I. Weiss. *Logo Recognition Using Geometric Invariants*. International Conference on Document Analysis and Recognition (ICDAR), pp. 894-897, 1993.

[6] M. Rusinol and J. Llados. *Logo Spotting by a Bag-of-words Approach for Document Categorization*. ICDAR, pp. 111-115, 2009.

[7] V.-P. Le, M. Visani, C.-D. Tran, J-M. Ogier. *Logo Spotting for Document Categorization.* International Conference on Pattern Recognition (ICPR), pp. 3484-3487, 2012.

[8] G. Zhu and D. Doerman. *Automatic Document Logo Detection*. ICDAR, pp. 864–868, 2007.

[9] M. Ester, H.-P. Kriegel, J. Sander and X. Xu. *Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. In E. Simoudis, J. Han, & U. M. Fayyad (Eds,), Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp. 226-231, 1996.

[10] Z. Li, M. Schulte-Austum and M. Neschen. *Fast Logo Detection and Recognition in Document Images*. ICPR, pp. 2716-2719, 2010.

[11] T.-A. Pham, M. Delalandre and S. Barrat. *A contour-based method for logo detection*.ICPR, pp. 718-722, 2011.

[12] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, Second Edition. Cambridge University Press, 2004.

[13] R. Jain and D. Doermann. *Logo Retrieval in Document Images*. 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 135-139, 2012.

[14] M. Rusinol and J. Llados. *Efficient Logo Retrieval Through Hashing Shape Context Descriptors*. DAS 2010, pp. 215-222, 2010.