Contemporary Engineering Sciences, Vol. 9, 2016, no. 28, 1349 - 1357
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2016.6688

Word Spotting in Handwritten Arabic Documents Using Bag-Of-Descriptors

Youssef Elfakir, Ghizlane Khaissidi, Mostafa Mrabti and Driss Chenouni

Laboratory of computing and interdisciplinary physics/ENS, Fes, Morocco

Mounim El Yacoubi

SAMOVAR, Telecom Sud Paris, CNRS, Université Paris-Saclay, France

Abstract

This paper presents a query-by-example word spotting in handwritten Arabic documents, based on Scale Invariant Feature Transform (SIFT), without using any text word or line segmentation approach, because any errors affect to the subsequent word representation. First the interest points are automatically extracted from the images using SIFT detector, then, we use SIFT descriptor to represent each interest point in the images. In the end, we represent the image's regions as histogram of visual words. The validate study is conducted under a series of controlled experiments on handwritten Arabic documents images.

Keywords: Word spotting, Handwritten Arabic documents, Scale Invariant Feature, bag of visual word, Indexation

1 Introduction

Old manuscripts are a part of the richest cultural heritage and legacy of civilizations. Repetitive manual manipulation of fragile documents should be avoided because as it could destroy them. Digitalization, therefore, is a convenient solution for the preservation of these manuscripts. Many digitization projects, which treat Latin scripts, has been developed such as manuscripts d'Oc and d'Oïl in the Vatican Library [1], Better Access to Manuscripts and Browsing of Images [2], etc. The conception of recognition systems for degraded handwritten Arabic document images knows today a great expansion and appears as a necessity in order to exploit the wealth of information contained in ancient manuscripts. This paper deals with the problem of query-by-example word spotting in handwritten

Arabic documents. This operation needs a lot of time and effort to do by manual inspection. Many existing architectures on word spotting based on text, word or line segmentation steps [3, 4] used in the recognition systems to facilitate the search, Roy et al. [5] use string matching of character primitives segment. However, any segmentation errors of the document affect the subsequent word representations and matching steps. This explains why research on word spotting and retrieval is oriented towards segmentation-free methods. Leydier et al. describe in [6] a word retrieval engine; this approach does not need any layout segmentation and makes use of features fitted to any type of alphabet. In [7], Gatos et al. present an approach applied to historical printed documents without requiring any previous block or word segmentation step, the proposed method based on image-block-descriptors and used at a template matching process. Rothacker et al. [8] propose to combine the Bag-of-visual-word representation with Hidden Markov Models in a patch-based segmentation-free framework in handwritten documents. In [9], Almazán et al. use query-by-example paradigm where the local patches described by a bag-of-visual-words model powered by Scale-invariant feature transform descriptors. Then, Spatial Pyramid Matching is use to overcome the problem of spatial representation (anagram).

In this paper, we addresses the search's problem for information in handwritten Arabic documents, without using any segmentation step, because it is not easy and a perfect segmentation is unfeasible, for this reason, the researchers of are motivated to move towards segmentation-free methods. We address the search problem by using a Bag of Visual Words (BoVW) powered by Scale-invariant feature transform (SIFT) descriptors. The proposed method was applied to handwritten Arabic document images from the Ibn Sina dataset [10] and other Handwritten Arabic documents. The obtained results are satisfactory in term of recognition rate.

The remainder of this paper organized as follows: In Section 2, we present the feature extraction process based on Bag-of-visual-word powered by Scale-Invariant Feature Transform descriptors for word spotting in handwritten Arabic document. In the section 3, we present the results of the proposed approach. Finally, conclusion and perspective is given in Section 4.

2 Feature Extraction

In this work, we present an unsupervised method for spotting and searching query (Fig .3), for this, we use the Scale-Invariant Transform Feature (SIFT) detector and descriptor, The SIFT detector extracts from the handwritten Arabic document images a collection of key points. The SIFT descriptor is used to descript this key points. Due to canonization, descriptors are invariant to translations, rotations and scaling. The approach of SIFT feature detection taken in our implementation is similar with the one taken by Lowe et. [11].

2.1 Scale space parameters

2.1.1 Creating the Difference of Gaussian Pyramid

The SIFT detector and descriptor are constructed from the Gaussian scale space of the image source I(x). The Gaussian scale space is the function:

Figure 1 Gradient's Pyramide: 4 octaves and 5 gradients

$$L(x, y, \sigma) = G(x, y, \sigma) * \mathrm{I}(x, y) \quad \text{with}$$

$$\mathrm{G}(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Then, the difference of Gaussian (DoG), D(x, y, σ), is calculated as the difference between two filtered images, one with k multiplied by scale of the other (Fig .1).

$$\mathrm{D}(x, y, \sigma) = L(x, y, \mathrm{k}\sigma) - L(x, y, \sigma)$$

Where $K = 2/S$ and s is the number of scale gradient

2.1.2 Extrema Detection

The goal of this stage is to find the extrema points in the DOG pyramid. To detect the local maxima and minima of D(x, y, σ), each point is compared with the pixels of all its 26 neighbors (Fig 2). If this value is the minimum or maximum, this point is an extrema. Then, we eliminate the key point that have low contrast or localized on an edge.

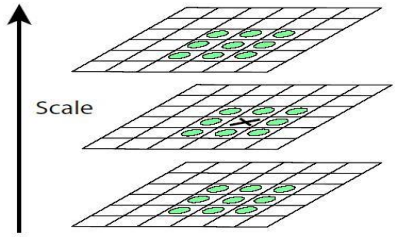

Figure 2: Sample point is selected only if it is a minimum or a maximum of these points

2.1.3 Orientation Assignment

In this stage, we assign a consistent orientation to the key points based on local image feature. An orientation histogram is formed from the gradient orientations of sample points within a region around the key point, for this, we calculate the gradient magnitude m(x, y), and orientation θ(x, y).

$$m(x,y) = \sqrt{(\mathrm{L(x+1,y)} - \mathrm{L(x-1,y)})^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{(L(x,y+1) - L(x,y-1)}{\mathrm{L(x+1,y)} - \mathrm{L(x-1,y)}}\right)$$

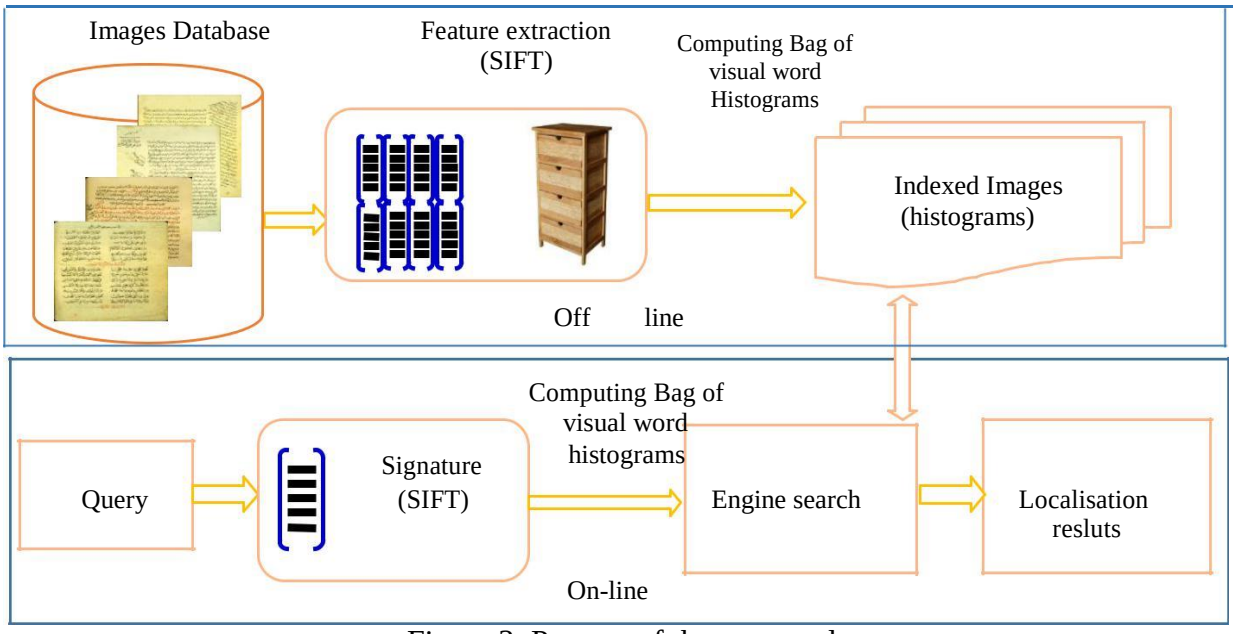

Figure 3: Process of the proposed system.

2.1.4 Descriptor Computation

In this stage, we compute the descriptor for each region in the image at each key point. Then, the gradient magnitudes and orientations sampled around the key point location. In our implementation, considering a region of 16 × 16 pixels around this key point, divided into 4 × 4 areas of 4 × 4 pixels, for each area is calculated a histogram of orientations with 8 bins. At each point of the area, the direction and magnitude of the gradient is calculated. Then the 16 histograms for with 8 bins are concatenated and normalized, and finally, provide the SIFT descriptor of each key point with 128 dimension (Fig .4).

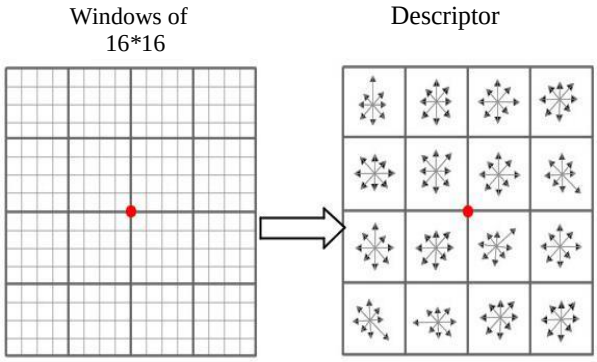

Figure 4: Descriptor Computation

2.2 *Image Representation with a Bag of Visual Words*

Bag of Visual Words (BoVW) is a popular technique for image classification inspired by models used in natural language processing based on a sparse histogram of occurrence counts of words.
The main steps are:

a) Extract features
b) Learn “visual vocabulary”
c) Quantize features using visual vocabulary
d) Represent image’s regions by frequencies of “ visual words”

In Handwritten Arabic documents, for H the height of a given query, approximately, we defined three widths, to synchronize it with queries of different sizes (Fig .5). Specifically, the geometry of the regions has been set to H*H, 2H*H and 3H*H. We divide the document images into a set of local regions, densely sampled; these local regions are the basic structure used to spot the words in the document. For a given query, we use the local regions to determine in the images the locations where the query has a greater probability to appear, the size of query should approximately equal to the size of the local region in the document. Then, the document images regions are represented by local feature representatives; using Scale-Invariant Feature Transform; around each interest point. In the second step, we represent the local regions by Histograms of Visual-Word. The most conveni-

ent region width for the matching step is determined at query time. Then, the SIFT descriptors are calculated and a codebook is used to quantize them into visual words by clustering the descriptor feature space into K different clusters by using the k-means algorithm. Afterwards, visual words are obtained by simply assigning to each SIFT descriptor the nearest code word of the codebook, each local region within the document is described by a histogram of accumulates frequencies $h_j = (h_j^1, h_j^2 .....h_j^k)$ where k is the number of the visual word and h_j^i is visual word occurrence of the region j. For matching, we use the nearest neighbor search (NNS).

Figure 5: Handwritten Arabic words in different size H*H, H*2H and H*3H

3 Results and Discussions

In this section, we present the result of the proposed approach for searching query in handwritten Arabic manuscripts. We use MATLAB to measure all score and running times of the different sections. We calculate the precision for different queries, as we see in (Fig. 6) and (Fig. 7), the precision and recall mean results depending on the codebook sizes.

$$\text{Precision} = \frac{|(\text{relevant document}) \cap (\text{retrieved document})|}{|(\text{retrieved document})|}$$

$$\text{Recall} = \frac{|(\text{relevant document}) \cap (\text{retrieved document})|}{|(\text{relevant document})|}$$

$$\text{mAP} = \frac{\sum_{q=1}^{Q} P(q)}{Q}$$

As can be seen in Table 1, the proposed approach provides a good performance at mAP. In order to evaluate the performance of the approach in handwritten Arabic documents, we change the sizes of codebook. (Fig .6) and (Fig .7), shows that the best mean average precision (81 %) is obtained for 100 codewords.

Table 1: Mean Average Precision and Recall for different codebook size

Codebook size	**80**	**100**	**150**	**200**
mean Precision	0,67	0,81	0,63	0,65
mean Recall	0,64	0,78	0,68	0,69

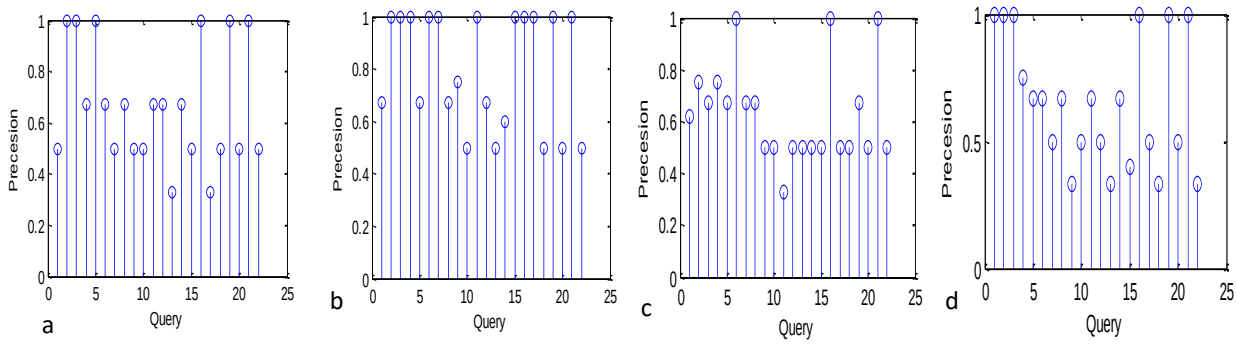

Figure 6: Precision at different codebook sizes: (a) 80, (b) 100, (c) 150 and (d) 200

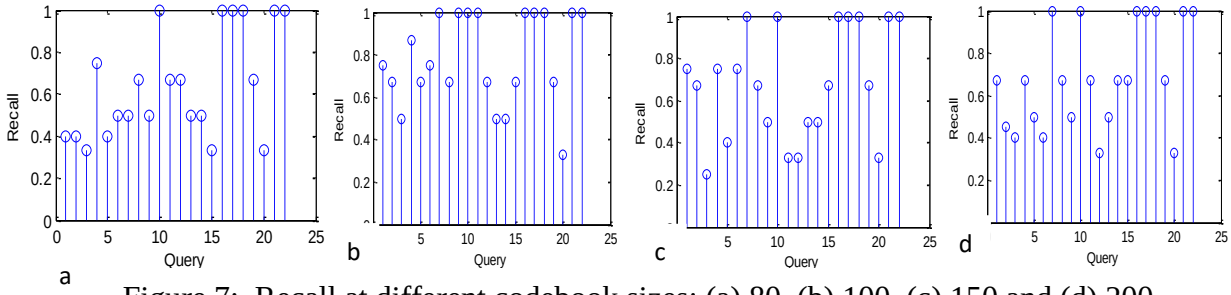

Figure 7: Recall at different codebook sizes: (a) 80, (b) 100, (c) 150 and (d) 200

"Table 2" shows the mean average precision of the approach proposed, and an ancient approach [12] applied to handwritten Arabic documents.

Table 2: Mean Average Precision

Method	**HOG+SVM[12]**	**Proposed approach**
mAP	0,68	0,81

4 Conclusion and perspective

In this paper, we have presented query-by-example word spotting in handwritten Arabic documents using SIFT algorithm combined with BoVW method, the proposed method was applied to handwritten Arabic document images from the Ibn Sina dataset and other Arabic documents. The obtained result (Table .1) is satisfactory in term of recognition rate. In future, this work can extend to enhance the performance of the system by adding some more relevant features.

References

[1] IRHT, coord. Maria Careri (Université de Chiet - membre associé à l'IRHT), Anne-Françoise Leurquin et Marie-Laure Savoye (2011-2021).

[2] Sylvie Calabretto, Andrea Bozzi, Jean-Marie Pinon, Numérisation des manuscrits médiévaux : le projet européen BAMBI, Lyon. décembre 1999.

[3] T. M. Rath and R. Manmatha. Word image matching using dynamic time warping, *Proceedings 2003 International Conference on Computer Vision and Pattern Recognition*, Vol 2, (2013), 521–527.
http://dx.doi.org/10.1109/cvpr.2003.1211511

[4] Y. Elfakir, G. Khaissidi, M. Mrabti, Z. Lakhliai, D. Chenouni, M. Elyacoubi, Contribution à l'indexation des documents manuscrits arabes scannés, *Mediterranean Telecommunication Journal*, **5** (2015), no. 2, 191-196.

[5] P. Roy, J. Ramel, N. Ragot, Word retrieval in historical document using character-primitives, *Proceedings of the International Conference on Document Analysis and Recognition*, (2011), 678–682.
http://dx.doi.org/10.1109/icdar.2011.142

[6] Y. Leydier, A. Ouji, F. Le Bourgeois, H. Emptoz, Towards an omnilingual word retrieval system for ancient manuscripts, *Pattern Recognition*, **42** (2009), no. 9, 2089–2105

[7] B. Gatos, I. Pratikakis, Segmentation-free word spotting in historical printed documents, *International Conference on Document Analysis and Recognition, Proceedings*, (2009), 271–275.
http://dx.doi.org/10.1109/icdar.2009.236

[8] L. Rothacker, M. Rusiñol, G. Fink, Bag-of-features HMMs for segmentation-free word spotting in handwritten documents, *12th International Conference on Document Analysis and Recognition, Proceedings*, (2013), 1305–1309. http://dx.doi.org/10.1109/icdar.2013.264

[9] J. Almazán, A. Gordo, A. Fornés, E. Valveny, Segmentation-free word spotting with exemplar SVMs, *Pattern Recognition*, **47** (2014), no. 12, 3967-3978. http://dx.doi.org/10.1016/j.patcog.2014.06.005

[10] R. F. Moghaddam, M. Cheriet, M. M. Adankon, K. Filonenko, and R. Wisnovsky, "IBN SINA: A database for research on processing and understanding of Arabic manuscripts images", *Proceedings of DAS'10 Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, (2010), Boston, MA, USA. http://dx.doi.org/10.1145/1815330.1815332

[11] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, **60** (2004), 91-110.
http://dx.doi.org/10.1023/b:visi.0000029664.99615.94

[12] G. Khaissidi, Y. Elfakir, M. Mrabti, D. Chenouni, Segmentation-free Word spotting for Handwritten Arabic Documents, *International Journal of Artificial*

Intelligence and Interactive Multimedia of Computer Applications, **4** (2016), no. 1, 6-10. http://dx.doi.org/10.9781/ijimai.2016.411

Received: June 21, 2016; Published: October 3, 2016