

The impact of the image processing in the indexation system

Youssef Elfakir, Ghizlane Khaissidi, Mostafa Mrabti, Driss Chenouni

Laboratory of Computing and Interdisciplinary Physics, ENS, Sidi Mohamed Ben Abdellah University, Morocco

Article Info

Article history:

Received Des 24, 2018

Revised Apr 15, 2019

Accepted Apr 28, 2019

Keywords:

Bag-of-visual word

Floating similarity distance

Handwritten arabic documents

Scale-invariant-feature transform

Word spotting

ABSTRACT

This paper presents an efficient word spotting system applied to handwritten Arabic documents, where images are represented with bag-of-visual-SIFT descriptors and a sliding window approach is used to locate the regions that are most similar to the query by following the query-by-example paragon. First, a pre-processing step is used to produce a better representation of the most informative features. Secondly, a region-based framework is deployed to represent each local region by a bag-of-visual-SIFT descriptors. Afterward, some experiments are in order to demonstrate the codebook size influence on the efficiency of the system, by analyzing the curse of dimensionality curve. In the end, to measure the similarity score, a floating distance based on the descriptor's number for each query is adopted. The experimental results prove the efficiency of the proposed processing steps in the word spotting system.

Copyright © 2019 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Elfakir Youssef,

Laboratory of Computing and Interdisciplinary Physics,

National Superior School,

University Sidi Mohammed Ben Abdellah, Fes, Morocco.

Email: youssef.elfakir1@usmba.ac.ma

1. INTRODUCTION

Old handwritten Arabic documents are part of the richest cultural heritage and contains a wealth of information. The repetitive manipulation of these manuscripts should be avoided as it could destroy them. To exploit this wealth of information contained in these manuscripts, digitalization is a convenient solution to preserve them. The recent advances in pattern recognition, storage, and network technology have paved the way for many digitization projects, which treat Latin scripts, such as manuscripts Better Access to Manuscripts and Browsing of Images [1], Electronic Access to Medieval Manuscripts (EAMMS) [2], etc.

This paper deals with the problem of word-by-example spotting in handwritten Arabic documents. From the survey of word spotting system, we found that a few researchers treat the handwritten Arabic documents, where a million documents had been writing in various disciplines. In the Arabic handwritten case, the recognition system is faced with various problems, which can be summarized as follows:

- Cursivity of the Arabic script
- Arabic language contains many diacritic marks
- Form of the same letter at the beginning and end of a word can be change
- People write with their own script

The word spotting process needs enough time and effort to be performed by manual inspection. To facilitate the search in numerical document images, numerous word spotting researchers had based on text line or word segmentation steps [3-6]. First, an initial step is performed to segment text into word candidates [7]. Then, candidates are represented by their sequences of features [8, 9]. In the end, to compare the query word and these candidates, a similarity measure based on Dynamic Time Warping [8] or Hidden Markov Model [10] is used. The main problem with these approaches is that they are very sensitives and need to perform a costly segmentation step to select candidate regions. When, the segmentation step is not usually easy and any error affects the representation's word, therefore the matching steps. This explains why

research on word spotting and retrieval is oriented towards segmentation-free methods over the last few years.

Late research on word spotting architecture has proposed approaches that do not need any segmentation step. In [11], Leydier et al. avoid the segmentation step in the word retrieval system, by using features fitted to any type of alphabet by computing local key points using a simple descriptor based on information's gradient. In the same approach, Zhang and Tan use features based on the Heat Kernel Signature [12]. The drawback of these proposed methods is that not scalable to large datasets, because the use a costly distance computation in the matching phase. In this way, Rothacker et al. propose in [13] to exploit the use of bag-of-visual-word representation with Hidden Markov Models to avoid segmentation step.

In [14], Almazán et al. avoid the segmentation step by representing documents with a grid of HOGs descriptors, where a sliding-window paragon is used to locate the locations that are most similar to the query in the dataset. Then, they use SVMs structure to get a better representation of the query. To solve the problem of memory, lately, the authors moved to use the concept of bag-of-feature representation. The method proposed by Marçal [15], use query-by-example [16] paradigm where the local patches are describing with a bag-of-visual-words model powered by Scale-invariant-feature-transform descriptors. Then, the spatial pyramid-matching framework is used.

In [17], Rodriguez uses a Model-based approach to measure the similarity between sequence's vector, where several features are extracted for all the images by using sliding window such as local gradient histogram [18], the zoning features [19] and the column features [20]. This sequence is mapped to a HMMs and a similarity measure is computed between them. In [21], Petitjean explains how the template matching influence in the context of pattern spotting in historical document images by integrating and evaluating different template matching methods.

The degradation on the handwritten document and can take different forms, such as the discoloration of ink, interfering patterns like ink bleed- through, show- through [22], etc. Therefore, before any process, as feature extraction or text segmentation, appropriate pre-processing is essential in order to correct the degradation [23]. In the present work, the document images are pre-processed, in order to enhance them and to eliminate the strongly interfering background, this step improve the extraction phase. To enable an efficient feature extraction, finding effective and robust features is an important task, which affects to the word retrieval performance [24]. In this case, the scale invariant feature transform algorithm (SIFT) has been applied to extract and to characterize interesting points in the document. This algorithm has shown their efficiency in previous research [25, 26]. To solve the problem of computer memory caused by descriptor's dimension, we propose to use bag-of-features approach [25], the SIFT descriptors have been used to create the histograms, and K-Means clustering has been applied for clustering to create the bag-of-visual-descriptors [27, 28]. Then, we represent the image's regions as histograms by using the bag of visual words [29, 30] method.

At this stage, and using data in high-dimensional spaces, the codebook size or the cluster's number became a very important task, which affects to the region's representation, subsequently, the matching phase. For this, we chose the best size of codebook by analyzing the curse of dimensionality curve [31, 32]. The last option is the histograms distance computation [33], a presentation of a proposed floating distance to measure the similarity score will follow.

The remainder of this paper is organized as follows. In Section 2, we first present the pre-processing stage to enhance the degradation on the handwritten document. Section 3 describes the proposed word spotting system. Afterward, Section 4 study the influence of the processing step in the proposed system where we report experimental results and analysis. Finally, conclusion and further research are drawn in section 5.

2. POST PRE-PROCESSING

The digitization of Arabic handwritten documents appears today as a necessity to preserve the integrity and rarity of space, However, the digitization is the first step in a process of classification and indexing to exploit all wealth of information. For this reason, we have adopted an indexing method for scanned Arabic handwritten documents. Images documents, and especially scanned handwritten documents, are complex and contain a large amount of relevant information. Most of this data is connected by relations of colors or intensities. Analysis and pre-processing of document images are avoided in some scenario of word spotting [14, 15], where the variation between colors or intensities in the document is not large. For this, and in order to overcome this existing variation in other documents like as Ibn Sina database we propose to start the indexing system by pre-processing step.

The text separation from image background is a very vast domain, where many research address this problem by a rough estimation of the text and background regions [34-37]. In [36], to identify the text and

background classes, a global binarization thresholding is used. Then, to adjust the threshold value, a noise model is building and used. In author work [23], and in order to identify the text, background, and uncertain pixels, a binarization paragon is presented, then, the uncertain pixels it binarized using a classifier trained based on the text and background classes.

In this paper, we address the enhancing and restoring problems of Arabic handwritten document images by using a series of multi-level classifiers [23]. We have modeled the pre-processing images step as shown in Figure 1 based on these classifiers, which can be used to an enhancement or restoration method. These multi-level classifiers are maps that extract relevant information for different levels: local, regional and global. And provide a value for each pixel in the image. There are several classifiers. In this work, we use the estimated background, the stroke gray level, and edge profile.

- Estimated background

Is a high-level classifier [38], use many other classifiers to arrive at an estimate background as near as possible to the true background of the image as shown in Figure 1-b.

- Stroke gray leve

This classifier provides a gray value for each pixel [23], the estimated intensities for a stroke is calculated by averaging the intensities of the pixels, an interpolated value will be assigned for the non-text pixels (background, figures interference, etc.).As shown in Figure 1-c.

- Edge profile

The edge profile is a classifier used to overcome the interference problem of the information. The calculation of the edge profile is based on the gradient of histogram in each region in the image as Figure1-d.

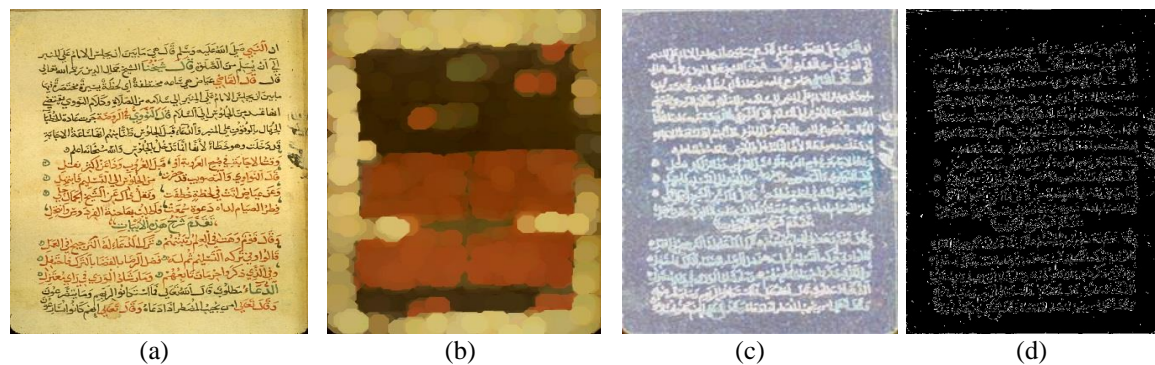


Figure 1. Pre-processing image steps, a) Original image, b) Estimated background, c) Stroke gray scale, d) Edge profile

Figure 2 shows the process of pre-processing step. We apply the restoration method to the handwritten Arabic documents. Figure 3 shows the pre-processing results:

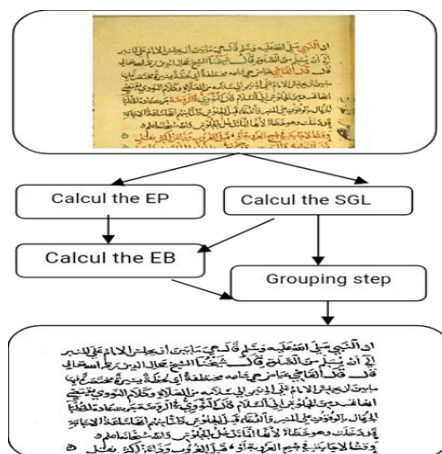


Figure 2. Pre-processing process

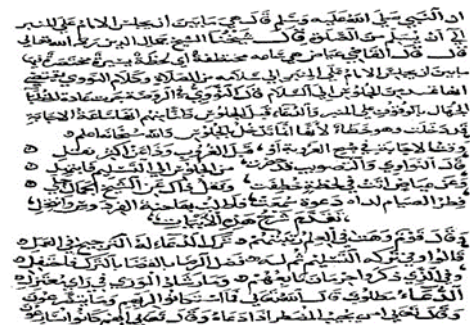


Figure 3. Pre-processed images

3. PROPOSED SYSTEM

We address the words spotting problem by using a Bag of Visual descriptors model powered by scale-invariant feature transform, which consists to describe each detected interest point. As we can see in [27], the performance of this model depends on the number of visual descriptors extracted from the image. In order to represent each region in the images by a histogram and taking account different words size, we densely divided the image into a set of local regions. For this, we define three widths $H \times H$, $2H \times H$ and $3H \times H$ to be synchronized with different locals region's sizes. The aim of this multi-scale representation is to capture all different words size as shown in Figure 4.

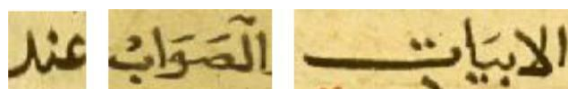


Figure 4. Handwritten Arabic words in different sizes $H \times H$, $H \times 2H$ and $H \times 3H$

The proposed method it has been applied to different handwritten document images. Figure 5 shows the process of the proposed word spotting system:

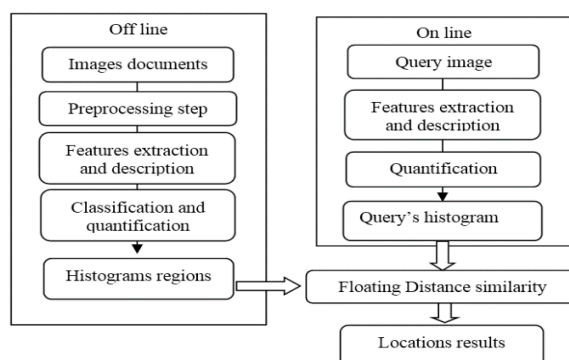


Figure 5. The process of the proposed word spotting system

3.1. Images representation

In [39], Lladós et al. showing the influence of word representations for handwritten words spotting in historical documents, and how a bag of visual words representation using SIFT descriptors can effectively perform the classical approaches, such as DTW based on sequence features. Here, we use the Scale-Invariant Transform Feature algorithm, due to canonization; descriptors are invariant to translations, rotations and scaling, and we show the impact of the pre-processing stage in the indexation system, then, we show how we can perform the results by using floating distance similarity. The Sift detector extracts the interest points from the images and then we describe them, the taken algorithm in our implementation is inspired by the one taken by Lowe et. [40].

The main drawback of this approach at this stage is that they use a costly distance computation and the need a great computer memory, which is not scalable to large datasets. The colossal number of detected key points in the document cause this problem. i.e. the average number of the key points at each region is 94, which is represented by a descriptor of 94×128 , each document image having in average more than 10 000 regions, in this case, we require approximately 114 MB of RAM to store each image. To solve this problem caused by dimension of these descriptors, instead to represent the image's regions by their descriptors we encoding each region by a histogram using a bag-of-visual-descriptors framework, which is inspired by models used in natural language processing, this technique is based on a sparse histogram of occurrence counts of visual words.

The main steps of this technique are:

- Features extraction
- Classification "codebook"
- Quantification
- Construct the region's histograms using codebook

In the second step, local regions are represented by using histograms of Bag-of-Visual-Descriptors. To achieve this representation, we use 10% of all descriptors to quantize them into K different clusters using the k-means algorithm. In the end, all regions are described with their histograms by assigning each descriptor in the region to the nearest visual descriptor in the codebook. So, each region is represented by a histogram of accumulates frequencies $H_{1,j} = H_{1,j}, H_{2,j}, H_{3,j}, H_{4,j}, H_{5,j}, \dots H_{N,j}$, where k represent the number of the visual descriptor in the codebook and $H_{N,j}$ represent the cumulative frequency of k visual descriptor in the j^{eme} region.

3.2. Images representation

The information extraction from handwritten document images is one of the major challenging topics in the field of document image analysis. In this part, we show how the pre-processing images step influence in the indexation system scenario. The word "يوم" written in three different ways in the same document as shown in Figure 6, which are written with different colors and different diacritical marks, while some regions in these words are degraded due to the antiquity of these manuscripts and the manual manipulation. For the reason above, and in order to compare the pre-processing impact, we have tested the proposed system in the grayscale and pre-treated images. We extract the interest points from each word, and for each one in the first word, we search the similar one in the second words as shown in Figure 7. As we see in Table 1, the number of the matched points in the grayscale image is high than the pre-treated image due to the faults detected. The most of this faults keys points are coming from non-text regions and does not describe the word trait in the image.



Figure 6. Word 'يوم' written in three different ways in the same document

Table 1. Number of the matched points

| Words matching | With preprocessing step | Without preprocessing step |
|---------------------------|-------------------------|----------------------------|
| 1 ^{eme} matching | 67 | 143 |
| 2 ^{eme} matching | 119 | 145 |
| 3 ^{eme} matching | 98 | 131 |



Figure 7. The matched points

To examine the influence of these faults detections and compare the results between pre-treated and no pre-treated images in the word-spotting framework, we apply the proposed system Figure 5. In the similarity step, we choose to fix threshold (Tf), and we return each distance similarity less of Tf. Then we calculate the test's accuracy of the system Fscore, which is a harmonic mean of precision and recall that means the ability of the system to provide all relevant solutions and reject others.

$$\text{The F-score is calculate as follows: } F_{\text{Score}} = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{recall} + \text{precision}}{\text{recall} + \text{precision}} \quad (1)$$

To evaluate the performance of the approach in handwritten Arabic documents, we change the size of codebook. As shown in Figure 8, the F-score results depending on the codebook size and the approach based on pre-processing provide a good performance in term of F-score Table 2, the best mean F-score (0,778) is obtained for 300 code words.

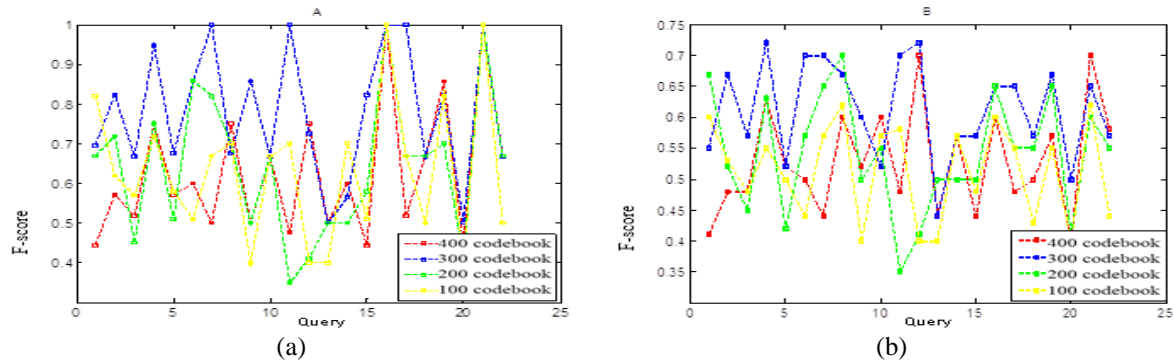


Figure 8. F-score at different codebook sizes, (a) pre-processed document, (b) no pre-processed document

Table 2. Number F-score at different codebook sizes

| Variable | 100 | 200 | 300 | 400 |
|---------------------------------------|-------|------|------|-------|
| F-Score for pre-processed document | 0,627 | 0,76 | 0,78 | 0,629 |
| F-Score for no pre-processed document | 0,53 | 0,61 | 0,63 | 0,51 |

3.3. The curse of dimensionality impact in word spotting system

In this section, we explain the curse of dimensionality impact in the bag-of-visual word system, this term was invented by Richard Bellman in [31, 32], the goal is to see diverse phenomena that appear when using data in high-dimensional spaces and analysing them. For this, we use two different Arabic handwritten documents from Gallica, which is the digital library of the French National Library, in open access. It brings digitized handwritten documents, magazines, images... First, we extract the interest points from each region in the pre-treated images using SIFT algorithm. Then, in the learning step, the descriptors of the 4 first images of the document are grouped to provide the cluster centres (codebook). In this stage, we use k-means algorithm with different number of k (centres), 100 to 900 centres. Then, to calculate the similarity between the histograms of each regions in the image's document and the query's histogram, we use the cos distance

$$\text{defined by: } S = 1 - \frac{\sum_{i=1}^N H_{i,j} R_i}{\sqrt{\sum_{i=1}^N H_{i,j}^2} \sqrt{\sum_{i=1}^N R_i^2}} \quad (2)$$

Where $H_{i,j}$ represent the occurrence of the i^{eme} centre of the codebook in the j^{eme} region, and $R(i)$ represent the occurrence of the i^{eme} centre of the codebook in the query. To judge that a region j is similar to the query, the cos distance S should be less than a certain threshold. For this, for each codebook size, we use various threshold between 0: 0.05: 0.7 and we calculate the recall and precision measures. The recall and precision curves show that the result depend on the threshold and codebook size as shown in Figure 9.

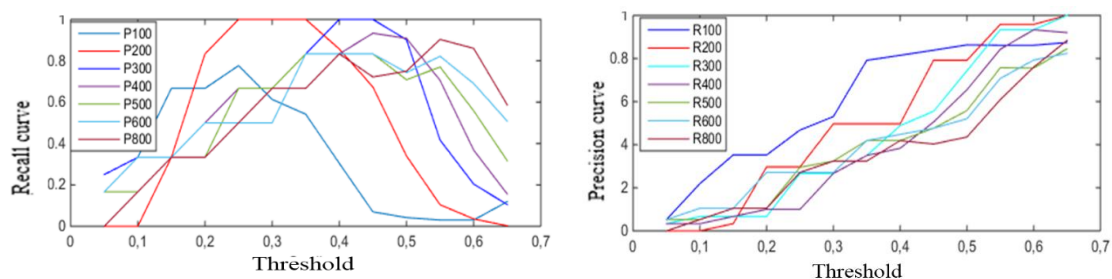


Figure 9. Recall and precision at different codebook sizes and several different threshold

To evaluate the impact of the codebook size, we calculate the F_{score} measure by using choosing the best threshold of each size as shown in Figure 10. We remark that the best result is given for $k=300$, the F_{score} decrease beyond this size due to the impact of the dimensional spaces.

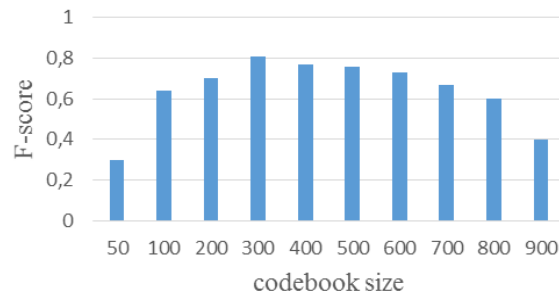


Figure 10. The F-sorce curve

Now, we search the influence of the codebook size on the threshold, from the (3), when the size increases (N), the probability to not find all visuals descriptors in the codebook for a given region's descriptors increase, that's mean the probability $P(R_i) = 0$ increase, and subsequently $P(H_{i,j} * R_i) = 0$ increase. Where R is the histogram of the query and H_j is the histogram of j^{eme} region in the document.

$$\text{So } D = \frac{\sum_{i=1}^N H_{i,j} R_i}{\sqrt{\sum_{i=1}^N H_{i,j}^2} \sqrt{\sum_{i=1}^N R_i^2}} \quad (3)$$

Decreased,

Thereafter, when the size N increases, the similarity distance S will increase ($S = 1 - D$).

$$N \nearrow \Rightarrow S \searrow \Rightarrow D \nearrow$$

Figure 11 shows the curve of the best threshold for each codebook size, as we see, the threshold depend on the size of the codebook. For this reason, and to overcome the problem linked to the curve of dimensionality, we use a codebook with 300 visual descriptors that give the best result.

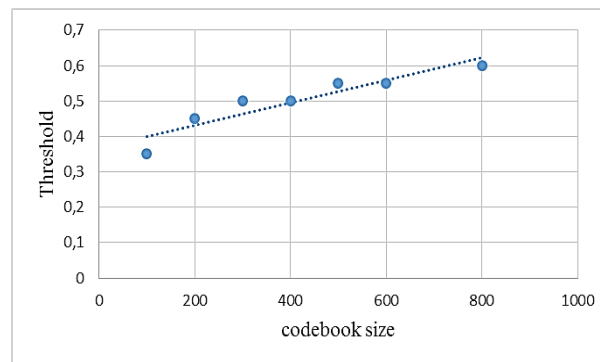


Figure 11. Best threshold for each codebook size

At this stage, we have demonstrate that the Fscore measure depend on the codebook size. To perform the experiments, we should search the impact of the descriptor's number (n) on the threshold. Form (3), when the number of interest points n increases, the probability to find all visuals descriptors for in histogram for a given region's descriptors increase, that mean probability of $P(R_i) \neq 0$ increase too. Thereafter, the probability $P(H_{i,j} * R_i) \neq 0$ increase. So D increases, with $0 \leq D \leq 1$. Thereafter, when n increases, the similarity distance S will decrease ($S = 1 - D$): $n \nearrow \Rightarrow S \nearrow \Rightarrow D \searrow$, which explain that the number of interest points influence on the threshold, or each word has a certain number of descriptors. For this, the distance similarity should be taken account the number of descriptors by using a floating threshold. Figure 12 represents the threshold curves according to the number of points of interest.

By analysing these curves, we can use the relation of the progression line as a floating threshold for each query. Or, if two different regions having even number of points of interest will have the same threshold, but different histograms, because they do not have the same interest points, and thereafter, different distances of similarity between these two regions and the query.

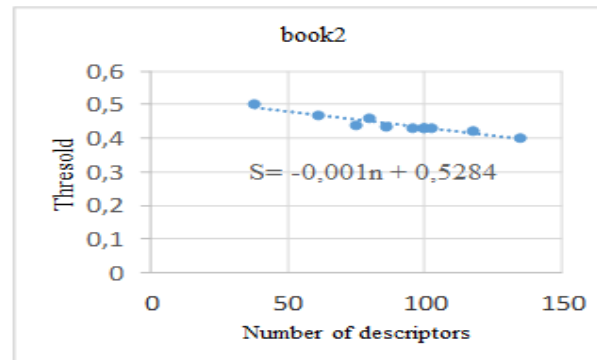


Figure 12. Threshold according to the number of descriptors

4. EXPERIMENTAL RESULTS

We tested our methodology on Arabic handwritten document from the digital library Gallica, Figure 13 presents qualitative results for two different documents. The used system is based on SIFT-BoVW descriptors with 300 visual words, and floating threshold. We report here some queries where the system yields automatically, which are similar to the query and without chose the K best similar results, which is a problem in other system [14, 41]. Then, we use a filtering step to select one best result when confusion regions are returning, based on their similarity scores and positions.

In comparison with state-of-the-art, we show the retrieval performance in terms of mAP as shown in Table 3, we can see how pre-processing by keeping the informative interest points in each word and discriminating the others. In addition, to overcome the problem linked to the various number of interest points in different regions by using floating threshold.



Figure 13. The retrieved images for some queries in the two evaluated documents

Table 3. Performance of the proposed method and other works

| Method | Precision |
|----------------------|-----------|
| Almazán et al. [14] | 68,4% |
| Proposed method | 83% |
| Howe et al. [42] | 79% |
| Liang et al.[43] | 67% |
| Fischer et al.[44] | 62% |
| Elfakir et al. [45] | 81% |
| Terasawa et al. [46] | 79% |

5. CONCLUSION

In this paper, we have presented an efficient free-segmentation word spotting approach for Arabic manuscripts. The proposed method present an excellent result when comparing other methods in literature. We have shown how the processing step can improve the result, in Bag-of-descriptors of SIFT system, by informative and discriminative features. Then, we have shown how we can improved the result by, first choosing the best size of codebook by analysing the curse of dimensionality curve, and secondly, the use of floating cos distance. Finally, we have presented a comparison with other methods, we tested our method using experimental setup based on MATLAB code applied to Gallica database

REFERENCES

- [1] Calabretto S., Bozzi A., Pinon J., "Digitization of medieval manuscripts (in French)," *the European project BAMBI, in: Proceedings of the conference Towards a new erudition: digitization and research in book history* (in French), Rencontres Jacques Cartier, Lyon, Dec 1999.
- [2] <http://www.hmm1.org/eamms/index.html>
- [3] Rath T.M., Manmatha R., "Word image matching using dynamic time warping," *Proceedings International Conference on Computer Vision and Pattern Recognition*, pp. 521-527, Feb 2013.
- [4] Elfakir Y., Khaissidi G., Mrabti M., Chenouni D., "Handwritten Arabic Documents Indexation using HOG Feature," *International Journal of Computer Applications*, vol. 126, no. 9, pp 14-18, Sep 2015.
- [5] Saeed K., Albakoor M., "Region growing based segmentation algorithm for typewritten and handwritten text recognition," *Applied Soft Computing*, vol. 9, pp. 608-617, 2009.
- [6] Louloudis G., Gatos B., Pratikakis I., *et al.*, "Text line and word segmentation of handwritten documents," *Pattern Recognition*, vol. 42, pp. 3169-3183, 2009.
- [7] Fitrianiingsih F., Madenda S., Ernastuti E., Widodo S., Rodiah R., "Cursive handwriting segmentation using ideal distance approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2863, 2017.
- [8] Rodriguez-Serrano J., Perronnin F., "A model-based sequence similarity with application to handwritten word-spotting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 2108-2120, 2012.
- [9] Marti U.-V., Bunke H., "Using a statistical language model to improve the performance of an HMM based cursive handwriting recognition systems," *Int. J. Pattern Recognit. Artif. Intell.*, pp. 65-90, 2001.
- [10] Griffiths Rath T., Manmatha R., "Word spotting for historical documents," *Int. J. Doc. Anal. Recognit.*, pp. 139-152, 2007.
- [11] Leydier Y., Ouji A., Le Bourgeois F., *et al.*, "Towards an omnilingual word retrieval system for ancient manuscripts," *Pattern Recognition*, vol. 42, no. 9, pp 2089-2105, 2009.
- [12] Zhang X., Tan C. L., "Segmentation-free key word spotting for handwritten documents based on heat kernel signature," in *International Conference on Document Analysis and Recognition*, pp. 827-831, 2013.
- [13] Rothacker L., Rusinol M., Fink G., "Bag-of-features HMMs for segmentation-free word spotting in handwritten documents," *12th International Conference on Document Analysis and Recognition, Proceedings*, pp. 1305-1309, 2013.
- [14] Almazán J., Gordo A., Fornés A., *et al.*, "Segmentation-free word spotting with exemplar SVMs," *Pattern Recognition*, vol. 47, no. 12, pp. 3967-3978, 2014.
- [15] Marcal M., Aldavert D., Toledo R., *et al.*, "Efficient segmentation free keyword spotting in historical document collections," *Pattern Recognition*, vol. 48, no. 2, pp. 545-555, Feb 2015.
- [16] Doungpaisan P. and Mingkhwan A., "Query by Example of Speaker Audio Signals using Power Spectrum and MFCCs," *International Journal of Electrical & Computer Engineering (IJECE)*, vol. 7, no. 6, pp. 2088-8708, 2017.
- [17] Rodriguez-Serrano J., Perronnin F. "A model-based sequence similarity with application to handwritten word-spotting," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 2108-2120, 2012.
- [18] Rodriguez JA., Perronnin F., "Local gradient histogram features for word spotting in unconstrained handwritten document," In *International conference on frontiers in handwriting recognition*, 2008.
- [19] Bunke H., Bengio S., Vinciarelli A., "Offline recognition of unconstrained handwritten texts using HMMs and statistical statistical language models," *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no. 6, pp. 709-720, 2004.
- [20] Marti U-V., Bunke H., "Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system," *Int J Pattern Recognit Artif Intell*, vol. 15, no. 1, pp. 65-90, 2001.
- [21] En S., Petitjean C., Nicolas S., *et al.*, "Pattern localization in historical document images via template matching," *23rd International Conference on Pattern Recognition (ICPR)*, Cancun, pp. 2054-2059, 2016.
- [22] Moghaddam R. F., Cheriet M., "Low quality document image modeling and enhancement," *International Journal on Document Analysis and Recognition*, vol. 11, no. 4, pp. 183-201, Mar 2009.
- [23] Moghaddam R. F., RSLDI M. C., "Restoration of single-sided low-quality document images," *Pattern Recognition*, vol. 42, no. 12, pp. 3355-3364, Dec 2009.
- [24] Marinai S., Miotti B., Soda G., "Digital Libraries and Document Image Retrieval Techniques: A Survey," *Learning Structure and Schemas from Documents, Springerlink*, no. 375, pp. 181-204, 2011.
- [25] Shekhar R., Jawahar C., "Word image retrieval using bag of visual words," In *Proceedings of the 10th IAPR International Workshop on, Document Analysis Systems (DAS)*, 2012.

- [26] Rothacker L., Rusinol M., Fink G., "Bag-of-features HMMs for segmentation-free word spotting in handwritten documents," In *Proceedings of the 21st International Conference on Document Analysis and Recognition (ICDAR)*, 2013.
- [27] Nowak E., Jurie F., Triggs B., "Sampling strategies for bag-of-features image classification," in *European Conference on Computer Vision, Lecture Notes in Computer Science(LNCS)*, vol. 3954, pp. 490-503, 2006.
- [28] Sankar K.P., Manmatha R., Jawahar C.V., "Large-scale document image retrieval by automatic word annotation," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 17, no. 1, pp. 1-17, 2013.
- [29] Lazebnik S., Schmid C., Ponce J., "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2169-2178, 2006.
- [30] Elfakir y., Khaissidi G., Mrabti M., *et al.*, "Bag-of-descriptors of SIFT for Segmentation-free word spotting in Handwritten Arabic documents," *Second International Conference on Natural Sciences and Technology in Manuscript Analysis, Proceedings*, Hamburg, Germany (ICNTMA) 2016.
- [31] Ernest Bellman R. *Dynamic programming*, Princeton University Press, Rand Corporation, 1957.
- [32] Ernest Bellman R., *Adaptive control processes: a guided tour*, Princeton University Press, 1961.
- [33] Jégou H., Douze M., Schmid C., "Product quantization for nearest neighbor search," *IEEETrans. Pattern Anal. Mach.Intell.*, vol. 33, no. 1, pp. 117-128, 2011.
- [34] Hedjam R., Farrahi Moghaddam R., Cheriet M., "A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images," *Pattern Recogn.*, vol. 44, pp. 2184-2196, 2011.
- [35] Chen Y., Leedham G., "Decompose algorithm for thresholding degraded historical document images," *IEE Proc. -Vis. Image Signal Process*, vol. 152, pp. 702-714, 2005.
- [36] Don H. S., "A noise attribute thresholding method for document image binarization," *Int. J. Document Anal. Recogn.*, vol. 4, pp. 131-138, 2001.
- [37] Shokri M., Tizhoosh H., "Q(k)-based image thresholding," in *CVR'04*, pp. 504-508, 2004.
- [38] Cheriet M., Farrahi Moghaddam R., Hedjam R., "A learning framework for the optimization and automation of document binarization methods," *CVIU*, vol. 117, pp. 269-280, 2013.
- [39] Lladós J., Rusiñol M., Fornée A., *et al.*, "On the influence of word representations for handwritten words potting in historical documents," *Int. J. Pattern Recognit. Artif. Intell.*, 26, 2012.
- [40] Lowe D. G., "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.
- [41] Rusiñol M., Aldavert D., Lladós R. J., "Efficient segmentation-free key word spotting in historical document collections," *Pattern Recognition*, vol. 48, pp. 545-555, 2015.
- [42] Howe H., Rath T., Ranmatha R., "Boosted decision trees for word recognition in handwritten document retrieval," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 377—383, 2005.
- [43] Liang Y., Fairhurst M., Guest R., "A synthesized word approach to word retrieval in handwritten documents," *Pattern Recognition*, vol. 45, no. 12, pp. 4225-4236, 2012.
- [44] Fischer A., Keller A., Frinken V., *et al.*, "Lexicon-free handwritten word spotting using character HMMs," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 934-942, 2012.
- [45] Elfakir Y., Khaissidi G., Mrabti M., Chenouni D., El Yacoubi M., "Word spotting in handwritten Arabic documents using bag-of-descriptors," *Contemporary Engineering Sciences*, vol. 9, no. 8, pp. 1349-1357, 2016.
- [46] Terasawa K., Tanaka Y., "Slit style HOG feature for document image word spotting," in *Proceeding of the International Conference on Document Analysis and Recognition*, pp. 116-120, 2009.