

A Word Spotting Method for Arabic Manuscripts Based on Speeded Up Robust Features Technique

Noureddine El Makhfi*

Faculty of Science and Technology, Department of Mathematics and Computer Science, PMIC Lab, 32003, Al-Hoceima, Morocco

ARTICLE INFO

Article history:

Received: 15 August, 2019

Accepted: 07 November, 2019

Online: 22 November, 2019

Keywords:

Arabic calligraphy

Arabic handwritten documents

Cultural heritage

Digital libraries

Features extraction

Text-Image format

Segmentation

Word matching

Word spotting

ABSTRACT

The diversity of manuscripts according to their contents, forms, organizations and presentations provides a data-rich structures. The aim is to disseminate this cultural heritage in the images format to the general public via digital libraries. However, handwriting is an obstacle to text recognition algorithms in images, especially cursive writing of Arabic calligraphy. Most current search engines used by digital libraries are based on metadata and structured data manually transcribed in Ascii format. In this article, we propose an original method of pattern recognition for searching the content of Arabic handwritten documents based on the Word Spotting technique. Our method is both effective and simple, it consists in extracting a set of features from the words we segment in the target images and comparing them with the features of the words in the requested images. The principle of the method is to characterize each word with the Speeded Up Robust Features algorithm whose goal is to find all occurrences of query words in the target image even in the case of low-resolution images. We tested our method on hundreds of pages of Arabic manuscripts from the National Library of Rabat and the Digital Library of Leipzig University. The results obtained are encouraging compared to other methods based on the same Word Spotting technique.

1. Introduction

This article is an extended version of an oral communication presented in 2019 at the 5th International Conference on Optimization and Applications (ICOA2019) [1].

Virtual libraries, developed by research institutions, are a means of preserving documentary heritage in the form of images. The old manuscripts preserved in these libraries are consulted by a minority of people because of the image content that is inaccessible by search engines. Some techniques used for indexing manuscript images include metadata, annotation and transcription.

Metadata provides a means of describing the content of manuscripts in image form. On the other hand, the transcription or annotation of manuscripts offers an alternative to partially deepen the search in the content of the images. However, these techniques are very limited in requiring a considerable effort to reproduce the content of images in text format (ASCII) and do not offer a thorough and complete search. We find that copiers use the same writing style in old manuscripts. Each word or element (letter, pseudo word, calligraphy, etc.) is repeated several times in the corpus of Arabic manuscripts. In the absence of photocopiers at

that time, most Arabic manuscripts are reproduced several times by the same copiers with the same styles of Arabic calligraphy (Naskh, Kufi, Diwani, Thuluth, Roqaa, etc.)

The idea is to reduce the number of elements to be transcribed manually when entering the content of images in Arabic manuscripts that use the same writing style. Our objective is to compare elements (content) of Arabic manuscripts based on a robust method. Comparing the content of manuscript images requires a multi-step process, each word or element being characterized by a unique signature. This signature must be invariable and must adapt to changes in scale, rotation, geometric variations and illumination. The extraction of the features of words or elements in the form of a signature, thus facilitates the automation of annotations or the complete transcription of Arabic manuscripts. Several research studies have been carried out in the field of access to the content of manuscripts in image form. However, most of this work is based on binary images, which results in a huge loss of information in the images. To overcome this problem, we used grayscale images to extract useful features in Arabic manuscript images. To this end, we have based ourselves on multi-scale theory using grayscale images to ensure maximum reliability when extracting all the features of the segmented words.

*Noureddine EL MAKHFI, Email: n.elmakhfi@gmail.com

In this paper, we present an original method for identifying the content of Arabic manuscript images. This method is based on the Word Spotting technique which allows the search for similar words in a target image to that of a query image. We used the Speeded Up Robust Features algorithm (SURF) [2]. This algorithm allows to extract the features of the images based on local invariant descriptors / detectors (at scale changes, rotation, luminosity variations, etc.). The SURF algorithm is known in an area of computer vision with its robustness and speed to detect and describe points of interest in 2D images. Arabic manuscripts presented in image form contain information that is repeated several times. The Word spotting technique we use allows us to search for similar words in the same manuscript in order to facilitate its identification, indexing or transcription.

2. Related Work

There are several recent projects, allowing the cataloguing and digitization of ancient manuscripts based on the metadata technique. We can mention the project called "Cultural Heritage" of the Digital Library of the Leipzig University [3]. This is a project aimed at the online dissemination of Arabic, Turkish and Persian Islamic manuscripts. The presentation of these manuscripts is in the form of high definition images, in different formats. Indexing and searching in the databases of this digital library are in Trilingual services (German, English, Arabic). Most of the current research work is focused on researching the content of manuscript images. Among the most commonly used methods, the Word Spotting technique which allows the identification of words in images. There are recent studies that address the problem of indexing manuscripts and the recognition of cursive writing, namely:

In several studies [4-9], the authors treat the Word Spotting technique as regions of interest composed of words in which each character follows its model. The indexing and recognition of words are done by a probability calculation of regions of interest formed by characters. Each candidate region of interest is classified using one of the following methods: Dynamic Time Warping (DTW) or Hidden Markov Models (HMM) [10]. Despite the possible changes in words caused by character models, the results are interesting. Another work [11] which concerns the search for words based on annotations in medieval manuscripts. The techniques used are the basis of the Neural Networks (NN) and Hidden Markov Models (HMM) methods. The major problem with this method is the manual aspect of the production of annotations or transcription.

Other works that deal with word search based on the measure of similarity. The words, images are characterized according to their shapes using generic descriptors [12]. Among the shape descriptors that is used for word detection we can mention Histogram of Oriented Gradients (HOG) [13]. This description provides a means to describe and detect characteristics of handwritten words in [14].

In recent works [15-19], word detection is based on the SIFT algorithm [20,21]. This SIFT descriptor uses points of interest to describe the shapes of the words to be detected. The main problem with this technique used for word detection is the execution time and the false matching of the interest points of the images. Despite the excellence of the method for characterizing image words, like

what is done in computer vision for the detection of 3D objects. The limitations of this method for word detection in 2D images can be improved by inserting other steps such as pre-processing.

In the general case, methods for detecting objects in images are represented by several research projects, particularly those in the field of image processing and computer vision. These methods are classified according to three approaches: contour, model-based and local extrema detection.

The first approach provides a means to detect storytellers based on the location of the shapes of the objects in the image. This approach exploits the geometric properties of objects in the image as corners to detect storyteller points. The second model-based approach provides a mapping of a theoretical model to points of interest in the intensity function. The latter approach provides a means to characterize and detect local extrema. Each extremum corresponds to points of strong intensity variations in the image.

2.1. Contour approach

The H. Moravec [22] detector is one of the older ones used for the interest points. This type of detector is highly dependent on containers and its major problem is its sensitivity to noise.

An advanced version of the H. Moravec detector by C. Harris and M. Stephen [23] is named the Harris detector. It's a corner detector. It is invariable to translation, to rotation. However, it does not provide good results in the case of change of illumination. Further improvements to the Harris detector were made by Mohr and Schmid [24]. P. Montesinos [25] has developed this detector to process even colour images. Other improvements were made by Y. Dufournaud [26] to make this detector invariable to changes in resolution. A highly advanced version of the Harris detector was developed by Mikolajczyk [27] to detect points of interest in scale space. This method consists of using Gaussian smoothing to get the points that have a local maximum in the scale space. A further improved version of the Harris detector that is invariable to affine transformations was developed by Harris-affine [28].

2.2. Theoretical signal model approach

The theoretical signal model approach was developed by Schmid [22]. It consists in bringing together shapes of objects in images with theoretical signal models to obtain more sub-pixellic precision. This approach allows to locate points in the shapes of complex objects and not only for corner detection. Rounded shapes such as circles and curves present an obstacle for corner detectors. Rohr [29, 30] developed a blur model based on the convolution of a Gaussian with a binary model to detect the junctions of lines. This model is based on several parameters for corner angles, namely the orientation of the symmetry axis and the opening. This model is also based on dot position, grayscale and blur. The adjustment of these parameters allows to obtain a theoretical signal approximating the observed signal in such a way as to obtain a result closer to the real. Corner detection is based on the least squares method in order to minimize parameters by offering a more precise model.

R. Deriche and T. Blaszk [31] optimized the processing time with the improvement of the Rohr method. The principle of their methods is based on the exponential function instead of using the Gaussian function. P. Brand and R. Mohr [32] improved

localization quality based on a theoretical model close to the signal by fine transformation. This reconciliation is performed with alignment tests, epipolar geometry calculation and 3D reconstruction. The authors announce that test results can achieve an accuracy of 0.1 pixels.

G. Giraudon and R. Deriche [31, 33, 34] proposed a theoretical corner model to increase detection accuracy. This model is used to track the behaviour of each detector. The authors demonstrated the relationship between responses in scale space and the position of the actual characteristic.

2.3. Approach based on local extrema

The approach based on local extrema allows to locate points of interest in the corresponding scale space. Several detectors have been processed in the literature and are widely used to locate local extrema. We quote some methods for detecting points of interest to characterize objects in the image:

T. Lindeberg [35] among the first who worked on scale space. The automatic scale detection method is used to characterize the points detected at the corresponding scale. The detected points are located at the maximum magnitude of the Laplacian. They correspond to the peak points of the scale characteristic of the associated local surface. This scale is determined in the main direction of dominant curvature.

The Scale Invariant Feature Transform (SIFT) method is presented by Lowe in [20, 21]. This method is based on the detection of points of interest corresponding to the local extrema. The principle of the method is based on the differences of the Gaussian DOG to detect invariable robust points in corresponding scales.

A study realised by Mikolajczyk and Schmid [36], proposed a method for the detection of invariable points of interest called Harris-Laplace. It is based on the former Harris detector, thus providing a localization of the local extrema. This method allows to keep only the local extrema at the neighbourhoods of the characteristic ladders in the form of invariable robust points of interest.

Another study realised by H. Bay et al [3], proposed a point of interest detection method based on the Speed Up Robust Features (SURF) algorithm. The method is also based on the location of local extrema as the other methods mentioned above: SIFT [21] by Lowe and Harris-Laplace [36] by Mikolajczyk. The authors of the SURF method have shown that their algorithm is more powerful, more robust. SURF's response time is largely optimized compared to the two SIFT and Harris-Laplace algorithms.

A recent study was developed by B. Bagasi et al. [37], It concerns the comparison between the SURF algorithm and the BRISK algorithm [38]. This is a study that deals with searching for images by content. It concerns the search for images by similarity in manuscript images without segmentation.

3. The Proposed Method

OCR optical character recognition software allows the conversion of characters in the image into the code (ASCII). These OCRs are only valid for printed text even though the recognition rate is not optimal for Arabic printed characters.

In this paper, we propose a method of searching the text in images of Arabic documents like that used in images of Latin manuscripts. In this context, due to the lack of methods for recognizing

manuscripts, it is difficult to access their content in image form. Annotation or transcription techniques require considerable effort when entering text equivalent to the content of manuscript images. For this reason, these techniques can be automated by adding the Word Spotting method. Our goal is to facilitate the identification of the words in each scanned manuscript in the form of images. It is a question of characterizing by a signature each handwritten element (word, pseudo-word, letter). This signature must be unique for words that are identical. The objective is to eliminate redundancies in repeating elements.

3.1. Principles of the proposed method

The method of identifying and searching for Arabic handwritten elements is illustrated in schematic form by a block diagram. We have an input target image containing one page of a manuscript and two types of input queries for which the proposed method uses the text query or image query containing the word to be searched. In the case, of the text entry that is common, we used a database in XML format containing the features of the words of each manuscript processed. In case the input is in the form of an image, we use the direct extraction of the features of the request image. In other words, our method is based on text-image matching for identification of the contents of the manuscripts. In the following we deal with the case of comparison of input query images with a target image.

After the feature comparison operations, the results are provided in a final step as an output, thus locating the detected words. The following figure shows a block diagram of the proposed Word spotting method:

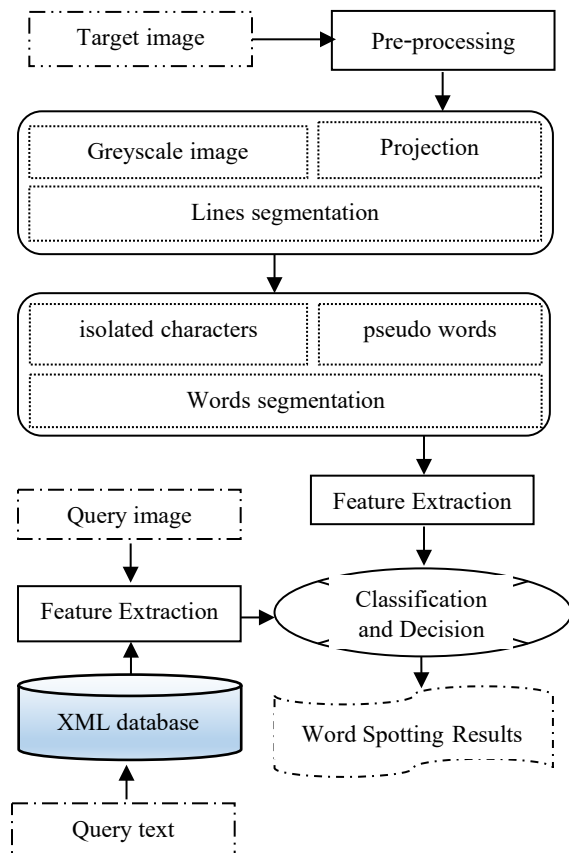


Figure 1: Synoptic diagram of the proposed Word Spotting method

• Acquisition and pre-processing

The first step is the acquisition of the manuscript in image form. Professional scanners make it possible to provide images with acceptable resolutions without damaging the manuscripts by the light emitted during scanning. Most often we obtain an image that contains two pages (left and right) in its raw state. We used as test images, manuscripts from the BNRM National Library in Rabat, Morocco. We also used images from a database of the Refaiya family [39] at the library of the University of Leipzig. The following figure shows the test image we used.



Figure 2: Original image [39]

After scanning, a series of pre-processing is required. Knowing that the scanned images are in their raw state in folders. Raw images take up a lot of space on storage disks. Image compression is necessary to reduce the storage capacity on the disk and speeds up processing time. Most of the professional scanners dedicated to manuscripts contain software and image processing algorithms. Other pre-processing must be applied to manuscript images such as: straightening, curvature correction, level spreading, contrast enhancement or detail enhancement, etc.



Figure 3: Left and right pages after pre-processing

In this article, we used the image on the right-hand side of the page. In most cases, Arabic manuscripts are written in different colours. We see in this example, that we have the colours of the red and black writing. The purpose of extracting text from the physical medium without losing important information. Colour segmentation of the image is an effective solution. It consists in selecting the pixel ranges in different colour layers. The most appropriate solution for our case is to segment each layer of the image separately and to assemble these layers with threshold values used. The HSV (Hue, Saturation, Value) space offers a simple way to detect thresholds. It allows easily distinguish intensity, saturation and hue.



Figure 4: Colour segmentation

• Line segmentation

The Word Spotting method requires a line segmentation [40] and subsequently a word segmentation to be able to locate the words to be searched in the image. In the first step, we will convert the colour image to a grayscale image whose purpose is to have a 1D signal. The projection algorithm [41] provides an effective way to facilitate line detection despite its sensitivity to noise. In this case, we use grayscale images instead of binary images because the projection of binary images [42] is very sensitive to line overlaps. It is widely used in the case of printed text because of the remarkable spacing between the lines. For our case, the projection of the grayscale image is represented by summing the intensity I at the coordinates of the image (x,y) to have the function of $f(y)$ according to the following relation:

$$f(y) = \sum_{x=0}^w I(x, y)$$

The function $f(y)$ represents the projected profile of image I . The calculation of the sum of pixels x along the y -axis is shown in Figure 5. Smoothing is essential to remove additional noise from the $f(y)$ function. The convolution of a Gaussian with this function allows to filter the high frequency noise as in [41].

$$p(y) = f(y) * g(y, \sigma)$$

$$g(y, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}$$

The grayscale image projection profile provides several useful information about the text in the image. It allows to illustrate the various characteristics like the basic lines (minima), the heights of the lines and the background of the image (maxima). In this case,

the texture effect is not considered because we have added a pre-processing step as a colour segmentation. The maxima noted on the curve represent the spacing between the lines. The maxima (local) can be calculated by applying the derivative of the p(y) function.

$$\frac{dp(y)}{dy} = 0$$

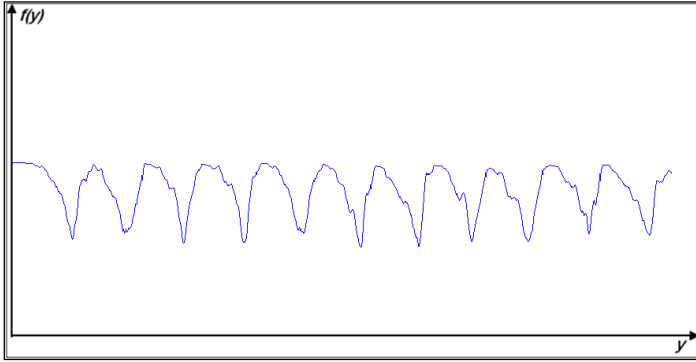


Figure 5: Line projection profile

The detection of local extrema facilitates the localization of the beginning and the end of lines. The following image shows the segmentation of the lines:



Figure 6: lines segmentation

• **Word segmentation**

The image in Figure 4, which is segmented in colour. It will be made binary by applying a global thresholding method. Among the thresholding methods, there is that of optimal Otsu thresholding [43] which we can apply thanks to the bimodal nature of the

histogram of this type of image. As an effect, the resulting image is affected by binary noise. As a solution, binary morphological filtering reduces the effect of this generated noise. We have already identified the text lines in the previous step of line segmentation. Dilation by applying a morphological filter of the binary image along each line allows the merging of pseudo-words and characters. The projection technique at the level of each line provides words. Overlapping characters prevent segmentation by this projection technique. The application of labelling of related components at the level of each line of the binary images makes it possible to locate isolated characters, pseudo-words and words. This method is also sensitive to noise, but it is effective. Additional noise from small areas is not considered by our algorithms. The following figure shows an example of segmentation of handwritten elements by applying labelling of related components (words, pseudo-words and characters).



Figure 7: Words and elements segmentation results

• **Feature extraction**

Arabic calligraphy has undergone several evolutions. It is considered an art of millennial history. The forms of Arabic letters follow refined rules. They change shape according to their position in the words (initial letter, median letter, final letter or isolated letter). Vowels can be designated by diacritical signs on words. Filtering of these diacritical signs is necessary to increase the quality of segmentation and recognition of words.

In most Arabic manuscripts, the writings follow uniform patterns according to the rules of the calligraphers. In many cases, these models are only the best-known fonts such as (Kufi, Andalus, Thuluth, Naskh, etc.). Words in manuscript images can be repeated several times with the same writing styles. The idea is to search for similar words in the same manuscript or in a collection of a calligrapher whose purpose is to use the occurrences of these words by our method.

Our objective is to choose such a method to extract the characteristics of words to be searched in manuscript images. We note that such a choice must be justified according to several constraints. In the literature, several approaches deal with recognition of forms. We opted for the use of invariable local detectors and descriptors to identify the forms of word writing. We justify the choice of these detectors by their acceptable response times and by their efficiency. On the other hand, we tested other methods, and could not obtain good results such as the theoretical signal model approach or the contour approach. We have the same character of the Arabic manuscript text that changes shape which makes the storyteller unstable. We opted for the approach based on local extrema because it uses points of interest recognized by their efficiency and by their optimised calculation times. In the literature, several approaches based on local extrema such as SURF descriptors offer a means of detecting these interest points.

In our case, the detection of words in images of Arabic manuscripts must take into consideration the styles of scripture. The approach using local extrema detection is based on variations in light intensity. However, in the case of small variations in light intensity, especially in manuscripts with almost identical intensities, the detection of word characteristics will be low. Another parameter to consider is the algorithmic cost and detector reliability. For example, the SIFT detector offers the interest points, with a higher computation time than the SURF detectors. That's why we use SURF's interest points to characterize the words in our Word Spotting method.

- **Classification and matching**

The degree of similarity of two words concerns the comparison of the characteristics of these two words. In our case, the characteristics of the words are based on SURF interest points. Comparison between points of interest requires a choice of the appropriate method to optimize processing time. Interest points of are characterized by their properties and their associated descriptor vectors. We compare these interest points in two stages: The first is to compare the trace sign of the Hessian matrix. This sign is from the Laplacian. It provides information on the meaning of blobs. Subsequently, two interest points to compare should have the same sign, if not different points.

The second is to compare the descriptor vectors of two interest points. Among the methods that can be used, the Euclidean distance or the Mahalanobis distance. The comparison is based on the calculation of distance and the correlation between these vectors.

4. Results and Discussion

Among the methods that are based on the Word Spotting technique we can mention the HADARA80P method [44] which

uses a dataset of 80 pages available on the internet [45] including images in Tiff format at high resolution. The documents handled are 9th century Arabic manuscripts. The authors of this method use three types of requests: by string, by polygon and by image which includes 25 test requests. Compared to the HADARA80P results, our proposed method provides satisfactory results even for low resolution images. However, the HADARA80P method requires high-quality images of about 300 dpi, especially when it comes to keyword images. In this case, images exceeding 50MB capacity can make their broadcasts difficult on the internet.

In our case, we used a microcomputer whose features is: CPU i5 Quad Core with a frequency of 2.4 GHz, 3 MB cache, 4 GB RAM and operating system Windows10 64-bit. We did tests on hundreds of pages and as well as on the dataset of HADARA80P. With twenty images request, we have achieved 95.27%-word matching accuracy with a run time of up to 910 ms for high memory capacity images.

Our method is to characterize the Arabic manuscripts. It identifies handwritten elements such as words, pseudo-words, isolated characters. The manuscripts we process are in the form of 2D images. The physical medium used in most manuscripts has a uniform texture and produces small variations in light intensity. However, ink generates large variations in light intensity. Indeed, the SURF points of interest used by our method characterise the text instead of the texture. Small variations in light intensity have a negligible effect on the detection of SURF interest points in manuscript images.

4.1. Word Matching

Word matching goes through several stages. These steps include extracting the features of words as interest points. Each word is represented by a set of interest points. These points are invariable to changes in scale, geometric variations, rotations and brightness variations. These invariances are shown by tests applied to the different words that are considered areas with a set of interest points. Regarding the effect of the texture of the physical medium was shown by the queries images 1, 2 and 4 on the table. In this case, we show as well that the texture allows to provide a negligible number of points. In addition, the accuracy of word matching depends on the number of interest points detected on each word. Many interest points allow to refine the description of each search word. To have high accuracy when matching words, it is necessary to use high quality images without increasing the execution time when searching for words.

The table shows the number of points detected on each query image. For this purpose, the interest points of the query images are located by small blue disks on these images. About the number of matching points between each query image and the target image is shown in the last column of the table.

The query images are compared with the target image with several of points detected in the target image equal to 773. We have a maximum of match points equal to 423 of the query image number 4 in the target image. However, the query image number 1 gives a minimum number of match points equal to 89.

The following figure shows all the query images used in the previous table. The target image is the one in the middle of the figure.

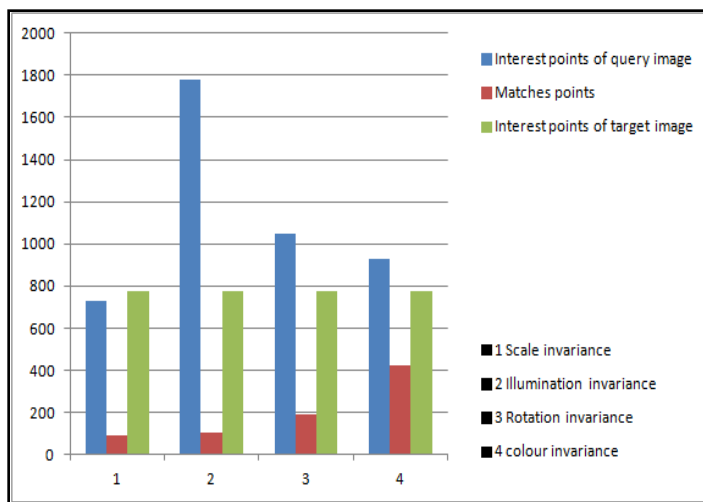
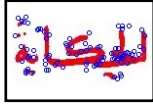
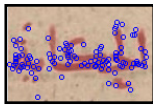
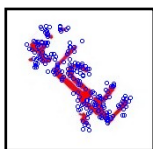
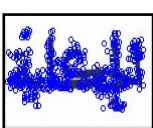


Figure 8: Histogram matching of interest points

Table 1: Experimental results of the matching points in query images with a target image

Query number	Invariance	Interest points	Query image	Matches points
1	Scale invariance	732		89
2	Illumination invariance	1782		107
3	Rotation invariance	1046		191
4	colour invariance	927		423

4.2. Application development

It is an application that is developed based on all the steps mentioned above (line segmentation, word and pseudo-word segmentation, feature extraction, classification). The extraction of the features is performed at the level of both query and target images. Word segmentation is an important tool to guide classification and identification. We used the SURF algorithm for the detection and description of features in the context of points of interest. This algorithm is known for its speed and robustness of its interest points in terms of invariance to scale changes, rotation, geometric variations and luminance. The following figure shows the interest points located in the target image and the query image on the left.

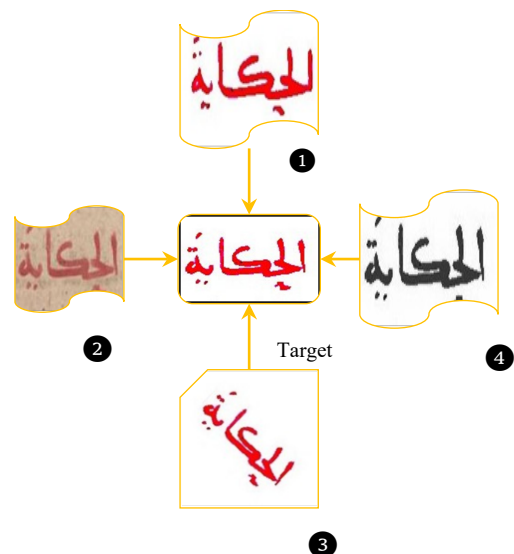


Figure 9: Four query images and the target image in the center



Figure 10: Interest points represented in the query images and in the target image

The objective is to develop a tool that allows access to the content of digitized Arabic manuscripts. The tools available for indexing handwritten documents in image form are very limited. In most cases, it is a search in databases describing the content of images with metadata. Our goal is to develop an application to identify text contained in images whose main interest is to integrate the Word Spotting technique for the search of occurrences of image word in handwritten documents. For this application, there are two possible cases:

- ✓ In a first case, the query image words are compared with those of the target image. Our method is based on interest points. However, an interest point belonging to a query image may belong to several words of the target image. Hence the need to segment the words of the target image and extract the features of each segmented word. Our interest is to search for all occurrences of a word image query in all pages of the same manuscript. The results of this research are acceptable despite some rejections due to segmentation problems in case of overlapping lines.
- ✓ In a second case, the query words can be text in ASCII format. This involves searching for the equivalent of text words in a database that stores the features of these words. Subsequently, the search for words in the target image is based on the features of the search words in the database.

The following figure shows the application we developed for the Word Spotting method. In the example, the comparison of the image query "history 'Al Hikayat' الحكياء" with a page image of an Arabic manuscript. The number of occurrences of this query image is equal to five times. As a perspective, we can expand this application to automate page annotations or help with transcription.

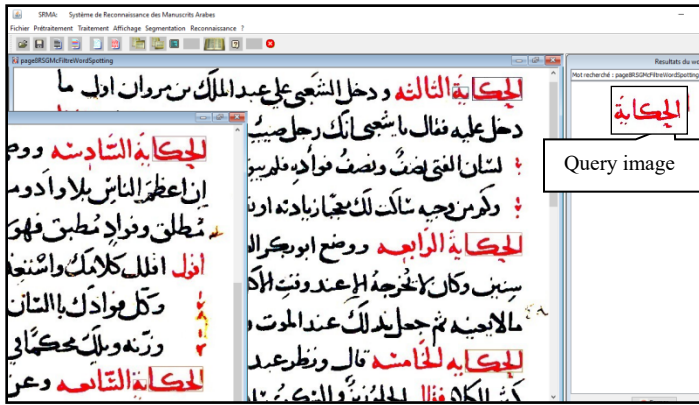


Figure 11: Word Spotting results

Conclusion

In this article, we have highlighted a method of searching for words in Arabic manuscripts. Most existing software (OCR) for Arabic text is only valid for printed text.

To this aim, we have proposed a word identification method based on the Word Spotting technique. The strength of this method lies in extracting the robust features of segmented words (or other elements: characters, pseudo-words). We use the invariant local detectors and descriptors of the SURF algorithm to extract the points of interest characterizing the words. The physical paper (or parchment) of the document has a negligible effect on the appearance of false points of interest belonging to the texture of the manuscript.

We have applied our method to hundreds of Arabic manuscripts from the National Library of Rabat and the Digital Library of Leipzig University. The test results of our method are excellent although some are very degraded. The limits of our method are in the case where the styles of writing and the fonts are varied, especially in the case the baldness of the lines where segregation is impossible.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] N. El makhfi, "Handwritten Arabic Word Spotting Using Speeded Up Robust Features Algorithm" in IEEE 5th International Conference on Optimization and Applications (ICOA), Kenitra, Morocco, 2019. <https://doi.org/10.1109/ICOA.2019.8727692>
- [2] Bay H., T. Tuytelaars, L.V. Gool, "SURF : Speeded Up Robust Features", in 9th European Conference on Computer Vision, Graz Austria, 404-417, 2006. <https://doi.org/10.1016/j.cviu.2007.09.014>
- [3] University of Leipzig. Project for the Cataloguing and Digitising of Islamic Manuscripts, www.islamic-manuscripts.net
- [4] R. Manmatha, C. Han, and E. M. Riseman, "Word spotting: A new approach to indexing handwriting", in Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 1996. <https://doi.org/10.1109/CVPR.1996.517139>
- [5] T. Rath, R. Manmatha, and V. Lavrenko, "A search engine for historical manuscript images", in Proceedings of the 27th annual international ACM

- SIGIR conference on Research and development in information retrieval (SIGIR), Sheffield, United Kingdom, 2004. <https://doi.org/10.1145/1008992.1009056>
- [6] T. Rath and R. Manmatha, "Word spotting for historical documents", IJDAR, 9 (2), 13-152, 2007. <http://dx.doi.org/10.1007/s10032-006-0035-8>
- [7] I. Yalniz and R. Manmatha, "An efficient framework for searching text in noisy documents", in DAS, Gold Cost, QLD, Australia, 2012. <https://doi.org/10.1109/DAS.2012.18>
- [8] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks", in IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 34 (2), 211 – 224, 2012. <https://doi.org/10.1109/TPAMI.2011.113>
- [9] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "HMM-based word spotting in handwritten documents using subword models", in ICPR, Istanbul, Turkey, 2010. <https://doi.org/10.1109/ICPR.2010.834>
- [10] F. Chen, L. Wilcox, and D. Bloomingberg, "Word spotting in scanned images using hidden markov models". In Acoustics, Speech, and Signal Processing, ICASSP-93., IEEE International Conference on, vol 5, 1-4, Minneapolis, MN, USA, 27-30 April 1993. <https://doi.org/10.1109/ICASSP.1993.319732>
- [11] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "Lexicon-free handwritten word spotting using character hmms", Pattern Recognition Letters, vol. 33, no. 7, pp. 934-942, 2012. <https://doi.org/10.1016/j.patrec.2011.09.009>
- [12] D. Zhang and G. Lu, "Review of shape representation and description techniques", PR, 37(1), 1-19, 2004. <https://doi.org/10.1016/j.patcog.2003.07.008>
- [13] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", CVPR, 2005. <https://doi.org/10.1109/CVPR.2005.177>
- [14] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Efficient Exemplar Word Spotting", in BMVC, 2012. [Online]. Available: <http://www.bmvc.org/bmvc/2012/BMVC/paper067/paper067.pdf>
- [15] S. Sudholt and G. A. Fink, "A Modified Isomap Approach to Manifold Learning in Word Spotting", in 37th German Conference on Pattern Recognition, ser. Lecture Notes in Computer Science, Aachen, Germany, 2015. https://link.springer.com/chapter/10.1007/978-3-319-24947-6_44
- [16] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Word Spotting and Recognition with Embedded Attributes" Transactions on Pattern Analysis and Machine Intelligence, 36 (12), pp. 2552-2566, 2014. <https://doi.org/10.1109/TPAMI.2014.2338414>
- [17] L. Rothacker and G. A. Fink, "Segmentation-free query-by-string word spotting with bag-of-features HMMs", in International Conference on Document Analysis and Recognition, Tunis, Tunisia, 2015. <https://doi.org/10.1109/ICDAR.2015.7333844>
- [18] M. Rusinol, D. Aldavert, R. Toledo, J. Lladós, "Efficient segmentation-free keyword spotting in historical document collections" Pattern Recognition, 48 (2), 545-555, 2015. <https://doi.org/10.1016/j.patcog.2014.08.021>
- [19] D. Aldavert, M. Rusinol, R. Toledo, and J. Lladós, "Integrating Visual and Textual Cues for Query-by-String Word Spotting", in International Conference on Document Analysis and Recognition, 511-515. Washington, DC, USA, 2013. <https://doi.org/10.1109/ICDAR.2013.108>
- [20] David. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". 60 (2), 91-110, 2004. [Online]. Available: <https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>
- [21] David G. Lowe, "Object Recognition from Local Scale-Invariant Features". Proc. of the International Conference on Computer Vision, Kerkyra, Greece, 1999. <https://doi.org/10.1109/ICCV.1999.790410>
- [22] Cordelia SCHMID, "Pairing of images by local gray-scale invariants, Application to the indexing of a base of objects", Ph.D thesis, National Polytechnic Institute of Grenoble, France, 1996.
- [23] C. Harris and M. Stephens, "A combined corner and edge detector", Proceedings of the 4th Alvey Vision Conference. 147-151, 1988. [Online]. Available: <http://www.bmva.org/bmvc/1988/avc-88-023.pdf>
- [24] Schmid, & R. Mohr, "Local grayvalue invariants for image retrieval", IEEE Transactions on Pattern Analysis and Machine Intelligence. 19 (5), 530-534, 1997. <https://doi.org/10.1109/34.589215>
- [25] P. Montesinos, V. Gouet, and R. Deriche, "Differential invariants for color images", in IEEE 40th International Conference on Pattern Recognition, Brisbane, Queensland, Australia, 1998. <https://doi.org/10.1109/ICPR.1998.711280>
- [26] Y. Dufournaud, Cordelia Schmid, Radu Horaud, "Matching Images with Different Resolutions", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hilton Head Island, SC, USA, 2000. <https://doi.org/10.1109/CVPR.2000.855876>
- [27] K. Mikolajczyk, C. Schmid, "Indexing based on scale invariant interest points", in IEEE 8th International Conference on Computer Vision (ICCV), Vancouver, BC, Canada, 2001. <https://doi.org/10.1109/ICCV.2001.937561>
- [28] K. Mikolajczyk, "Detection of local features invariant to affine transformation", Ph.D thesis, institut national polytechnique, Grenoble, France, 2002.

- [29] K. Rohr, "Recognizing corners by fitting parametric". *International Journal of Computer Vision*, **9**(3): 213-230, 1992. [Online]. Available: <https://link.springer.com/article/10.1007/BF00133702>
- [30] k. Rohr, "Über die Modellierung und Identifikation charakteristischer Grauwertverläufe in Realweltbildern", In 12 DAGM-Symposium Mustererkennung, 217-224, 1990. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-84305-1_25
- [31] R. Deriche & T. Blaska, "Recovering and characterizing image feature using an efficient model based approach". In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 1993. <https://doi.org/10.1109/CVPR.1993.341079>
- [32] P. Brand & R. Mohr, "Accuracy in image measure". in Proceedings of the SPIE Conference on Videometrics III, **2350**, 218-228, Boston, MA, United States, 1994. <https://doi.org/10.1117/12.189134>
- [33] R. Deriche & G. Giraudon, "Accurate corner detection: an analytical study". In Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan, 1990. <https://doi.org/10.1109/ICCV.1990.139495>
- [34] R. Deriche & G. Giraudon, "A computational approach for corner and vertex detection". *International Journal of Computer Vision*, **10**(2): 101-124, 1993. [Online]. Available: <https://link.springer.com/article/10.1007/BF01420733>
- [35] T. lindeberg, "Feature detection with automatic scale selection", *International Journal of Computer Vision*, **30** (2), 79–116, 1998. [Online]. Available: <https://link.springer.com/article/10.1023/A:1008045108935>
- [36] K. Mikolajczyk, & C. Schmid, "A performance evaluation of local descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI**, **27** (10), 1615 – 1630, 2005. [Online]. Available: <https://doi.org/10.1109/TPAMI.2005.188>
- [37] B. Bagasi, L. A. Elrefaei, "Arabic Manuscript Content Based Image Retrieval: A Comparison between SURF and BRISK Local Features", *International Journal of Computing and Digital Systems* ISSN (2210-142X) *Int. J. Com. Dig. Sys.* **7** (6), 355-364, 2018. [Online]. Available: <http://dx.doi.org/10.12785/ijcds/070604>
- [38] S. Leutenegger, M. Chili, and RY. Siegwart, "BRISK: Binary Robust invariant scalable keypoints" in 2011 IEEE International Conference. *Computer Vision (ICCV)*, 2548-2555, 2011, Barcelona, Spain <https://doi.org/10.1109/ICCV.2011.6126542>
- [39] The Refaiya Family Library, Leipzig University Library. <https://www.refaiya.uni-leipzig.de/content/index.xml?lang=en>
- [40] N. El makhfi, O. El bannay. "SCALE-SPACE APPROACH FOR CHARACTER SEGMENTATION IN SCANNED IMAGES OF ARABIC DOCUMENTS". *Journal of Theoretical and Applied Information Technology*, **94** (1), 2016. [Online]. Available: <http://www.jatit.org/volumes/Vol94No2/19Vol94No2.pdf>
- [41] R. Manmatha, N. Srimal, "Scale space technique for word segmentation in handwritten manuscripts", In 2nd International Conference on Scale Space Theory in Computer Vision (Scale-Space 99). **1682** (2), 22-33, Corfu, GRECE, 1999. [Online]. Available: <http://ciir.cs.umass.edu/pubfiles/mm-27.pdf>
- [42] J. Ha, R. M. Haralick, and I. T. Phillips. "Document page decomposition by the bounding-box projection technique". In Proceedings of the Third International Conference on Document Analysis and Recognition (ICDAR), pages 1119 1122, Montreal, Quebec, Canada, 1995. [Online]. Available: <https://pdfs.semanticscholar.org/e896/99bf5a1593490d6d93699ca17f3c75434178.pdf>
- [43] N. Otsu, "A threshold selection method from greyscale histogram", in *IEEE Transactions on Systems, Man, and Cybernetics*. on *SMC*, **9** (1), 62-66, 1979. <https://doi.org/10.1109/TSMC.1979.4310076>
- [44] W. Pantke, M. Dennhardt, D. Fecker, V. Möargner, T. Fingscheidt, "An Historical Handwritten Arabic Dataset for Segmentation-Free Word Spotting -HADARA80P", in proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICHFR), 15-20, Heraklion, Greece, 2014. <https://doi.org/10.1109/ICFHR.2014.11>
- [45] Technische Universität Braunschweig, HADARA80P Dataset. [Online]. Available: <https://www.ifn.ing.tu-bs.de/research/data/HADARA80P>