



HAL
open science

Use of SLIC superpixels for ancient document image enhancement and segmentation

Maroua Mehri, Nabil Sliti, Pierre Héroux, Petra Gomez-Krämer, Najoua Essoukri Ben Amara, Rémy Mullot

► **To cite this version:**

Maroua Mehri, Nabil Sliti, Pierre Héroux, Petra Gomez-Krämer, Najoua Essoukri Ben Amara, et al.. Use of SLIC superpixels for ancient document image enhancement and segmentation. Document Recognition and Retrieval, Feb 2015, San Francisco, United States. 10.1117/12.2076020 . hal-01119158

HAL Id: hal-01119158

<https://hal.science/hal-01119158>

Submitted on 23 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Use of SLIC superpixels for ancient document image enhancement and segmentation

Maroua Mehri^{a,b}, Nabil Sliti^c, Pierre Héroux^b, Petra Gomez-Krämer^a,
Najoua Essoukri Ben Amara^c, and Rémy Mullot^a

^aL3i Laboratory, University of La Rochelle, La Rochelle, France

^bLITIS Laboratory, University of Rouen, Saint-Etienne-du-Rouvray, France

^cSAGE Laboratory, University of Sousse, Sousse, Tunisia

ABSTRACT

Designing reliable and fast segmentation algorithms of ancient documents has been a topic of major interest for many libraries and the prime issue of research in the document analysis community. Thus, we propose in this article a fast ancient document enhancement and segmentation algorithm based on using Simple Linear Iterative Clustering (SLIC) superpixels and Gabor descriptors in a multi-scale approach. Firstly, in order to obtain enhanced backgrounds of noisy ancient documents, a novel foreground/background segmentation algorithm based on SLIC superpixels, is introduced. Once, the SLIC technique is carried out, the background and foreground superpixels are classified. Then, an enhanced and non-noisy background is achieved after processing the background superpixels. Subsequently, Gabor descriptors are only extracted from the selected foreground superpixels of the enhanced gray-level ancient book document images by adopting a multi-scale approach. Finally, for ancient document image segmentation, a foreground superpixel clustering task is performed by partitioning Gabor-based feature sets into compact and well-separated clusters in the feature space. The proposed algorithm does not assume any *a priori* information regarding document image content and structure and provides interesting results on a large corpus of ancient documents. Qualitative and numerical experiments are given to demonstrate the enhancement and segmentation quality.

Keywords: Ancient document images, enhancement, segmentation, SLIC superpixels, Gabor features, multi-scale approach.

1. INTRODUCTION

Since the early 1990s, libraries and museums have conducted large digitization campaigns with cultural heritage documents and scientific resources for ensuring a restoration and lasting preservation of historical collections and promoting worldwide accessibility to cultural patrimony which requires to be protected from further deterioration and damages caused by repetitive handling [1]. Due to the huge amount of numeric high-quality reproductions induced by the rapid growth of digital libraries worldwide, many challenges and open issues have been raised, such as providing a reliable document interpretation system and developing an efficient content-based image retrieval tool. Thus, a raising interest to ancient document image analysis has been generated since it helps to reach the objective of ensuring a relevant indexing and retrieval of digitized resources and offering a structured access to large sets of cultural heritage documents [2]. The ancient document image analysis starts by segmenting a document in order to find and classify homogeneous regions or zones, such as graphic and textual regions [3]. Finding graphic regions can be used to segment and analyze the graphical part in historical heritage, such as the drop caps (*i.e.*, the first letters at the beginning of a paragraph) [4] while determining text zones can be used as a pre-processing stage for character recognition [5], text line extraction [6] and handwriting recognition [7], *etc.* In this work, we are interested in ancient document image analysis. Our goal consists of identifying homogenous regions or similar groups of pixels sharing some visual characteristics by labeling and grouping pixels from ancient document images.

E-mails: {maroua.mehri, petra.gomez, remy.mullot}@univ-lr.fr, pierre.heroux@univ-rouen.fr, selliti.nabil@gmail.com and najoua.benamara@eniso.rnu.tn

Recently, a family of pixel-based algorithms for document image content and region extraction in document images has been proposed [8, 9, 10]. The document image analysis issue had been considered as texture segmentation and classification [3]. Kise precised that the most relevant methods used to analyze pages with overlapping or unconstrained layouts are based on signal properties of page components by investigating texture-based features and techniques [11]. Thus, in this article a texture-based approach is proposed in order to assist the document image analysis without hypothesis on the document structure, neither on the document model nor the typographical parameters. To our knowledge, there are a very limited number of comparative studies of texture-based approaches in the field of ancient document characterization and segmentation. For instance, Mehri *et al.* compared some of the well-known texture-based approaches: the auto-correlation function [12], Gray Level Co-occurrence Matrix (GLCM) [13], and Gabor filters [14], used for segmentation of digitized simplified ancient document images [15]. Qualitative and quantitative results were obtained on simplified ancient document images. They concluded that Gabor features performed better than auto-correlation and GLCM ones for font segmentation on the one hand, and for distinguishing textual regions from graphical ones on the other hand. Therefore, Gabor features are investigated in this article for ancient document image analysis.

Nevertheless, the literature shows that many different problems may arise related to the particularities of ancient documents such as a large variability of the page layout: noise and degradation caused by copying, scanning and aging (staining, mold or moisture, and the faded out ink and uneven lighting due to the folded and corrugated parchment or papyrus, *etc.*), page skew, complicated layout, random alignment, specific fonts, presence of embellishments, irregular spacings between the characters, words, lines, paragraphs and margins, overlapping object boundaries, varying text column widths, interspersed graphics, text in different orientations, presence of black borders, and superimposition of information layers (stamps, handwritten notes, noise, back-to-front interference or ink that was bleeding through, *etc.*). These characteristics complicate the characterization or segmentation of ancient documents and make the processing of this kind of document a difficult task [16, 2]. Thus, a pre-processing phase is necessary to remove background noise (*e.g.* superfluous information appearing from the verso, *etc.*). Indeed, in this work a novel foreground-background segmentation algorithm is introduced before the Gabor feature extraction and analysis task to obtain enhanced backgrounds of noisy ancient documents. The proposed algorithm is based on Simple Linear Iterative Clustering (SLIC) superpixels. Therefore, a fast automatic ancient document enhancement and segmentation algorithm based on using SLIC superpixels and Gabor descriptors in a multi-scale approach is proposed in this article.

The remainder of this article is organized as follows: in Section 2, the superpixel technique used for document image enhancement is described briefly. Section 3 details different Gabor-based approaches proposed in the literature and presents the investigated Gabor features. Section 4 presents the proposed ancient document enhancement and segmentation algorithm. In Section 5, we outline the experimental protocol by describing the experimental corpus and ground-truth. In addition, we discuss the pertinence of experiments by computing a classification metric for an evaluation of accuracy. Qualitative results are also given to demonstrate the enhancement and segmentation quality. Our conclusions and future work are presented in Section 6.

2. SUPERPIXEL TECHNIQUE

The superpixel technique has recently been used and has increasingly been applied in many computer vision applications, such as Gaussian noise estimation [17], class segmentation and object localization [18], *etc.* Instead of using a rigid structure of pixel grid for feature extraction and processing at the pixel level for segmentation, localization, and classification issues, the superpixel technique has been used as the basic unit and become a consistent alternative. The superpixel approach has the advantage to be faster, more memory efficient, and more interesting to compute image features on each superpixel center than on each image pixel [19]. It consists of grouping pixels sharing similar characteristics or properties (*e.g.* texture cues, contour, color, *etc.*) into a significant polygon-shaped region. Indeed, the superpixel approach produces an over-segmented image representing a compact content map [19].

Recently, the superpixel approach has gained great attention of many researchers in document image analysis fields. For instance, Cohen *et al.* separated drawings from background and noise of ancient documents by using spatial and color features which were extracted from superpixels [20]. Mehri *et al.* used an entropy rate superpixel segmentation algorithm based on a novel objective function [21] to obtain enhanced backgrounds of

noisy ancient documents [22]. By setting the number of superpixels equal to 5% of image pixels, the entropy rate superpixel segmentation algorithm is carried out on ancient document image. Then, the background and foreground superpixels are classified. Subsequently, an enhanced and non-noisy background is achieved after processing the background superpixels, *i.e.* the background superpixels and the pixels belonging to them have a value of white pixel (*i.e.* a 255 gray-level value).

Many superpixel algorithms have been proposed in the literature. Achanta *et al.* classified the existing superpixel methods into three classes: the graph-based, gradient-ascent-based, and SLIC superpixels [19]. The graph-based approaches consider each image pixel as a node in a graph. Then, the superpixels are generated by minimizing a cost function based on a the obtained graph. The gradient-ascent-based approaches process by iteratively refine the classification of samples on homogeneous and compact clusters until a convergence criterion is achieved to form superpixels. Finally, the SLIC approach is based on grouping pixels based on their color similarity and proximity in the image plane, by considering the five-dimensional space $[l, a, b, x, y]$, where $\{l, a, b\}$ represent the three pixel coordinates in the CIELAB color space, and $\{x, y\}$ are the pixel spatial coordinates.

The SLIC technique is based on an adaptive form of k-means clustering approach to well produce superpixels. By initiating the desired number of superpixels k_s , the SLIC algorithm defines the k_s initial cluster centers $C_i = [l_i, a_i, b_i, x_i, y_i]^T$, where $0 \leq i \leq k_s$, which form a structure of regular equally-sized superpixel grid. In order to generate superpixels with similar sizes, the grid interval S is set to $\sqrt{\frac{N}{k_s}}$, where N represents the number of pixels. Then, each pixel is assigned to the nearest cluster center. Finally, an update task is performed to control the cluster center according to the mean $[l, a, b, x, y]^T$ vector of all the pixels belonging to the cluster. Achanta *et al.* suggested to repeat 10 times the assignment and update steps of the SLIC approach to ensure good segmentation results [19]. They stated the SLIC technique as the best superpixel method comparing to five state-of-the-art superpixel methods in terms of the adherence to boundaries, computational and memory efficiency, segmentation performance, ease of use, parameter tuning, ability to control superpixel compactness, regularity and number, *etc.*

3. GABOR FEATURES

The investigated texture features tested in our work are Gabor descriptors which are extracted using the multi-channel Gabor filtering technique. The original Gabor elementary functions were firstly proposed by Gabor [14]. The multi-channel Gabor filtering was inspired by the multi-channel filtering theory first proposed by Campbell and Robson for processing of visual information of the human visual system [23]. Daugman modeled the processing of visual information of the human visual system by the two-dimensional multi-channel Gabor functions, which are local spatial bandpass filters [24]. The main idea of the multi-channel filtering technique is to exploit the differences in dominant sizes and orientations of different textures by decomposing the original image into several filtered images with limited spectral information. The two-dimensional Gabor functions have the advantage to have conjoint resolution information in both the two-dimensional spatial and Fourier domains. The filtered images are proceeded by tuning the analyzed image to combinations of frequency and orientation in a narrow range, which are referred to channels and interpreted as band-pass filters. By applying a bank of Gabor filters, the specified channels cover the spatial-frequency domain. Ursani *et al.* presented an empirical comparison between texture features based on the discrete Fourier transform and Gabor filters for texture recognition and retrieval [25]. They proved that analyzing Gabor features in image datasets containing noisy and rotated variants of texture performed better than analyzing Fourier descriptors for texture recognition and retrieval. Hence, Gabor filters have been shown to have good performance, due to its optimal localization properties to capture information in both spatial and frequency domains from the analyzed images, as opposed to Fourier transform.

A two-dimensional Gabor filter is a linear selective band-pass filter, dependent on two parameters: spatial frequency f_g and orientation θ_g , which characterize the specified channel. It consists of a Gaussian kernel function modulated by a sinusoidal plane wave. The spatial frequency f determines the distance from the Gaussian centers to the origin while the orientation θ_g specifies the angle from the horizontal axis, *i.e.* α -axis to the Gaussian centers. The multi-channel Gabor filtering approach is inherently multi-resolutional, which is a close relative of the wavelet transform [26].

The Gabor transform of an image $I(x, y)$ is:

$$I_{G(f_g, \theta_g)}(x, y) = \sum_{\alpha \in \Omega} \sum_{\beta \in \Omega} I(x + \alpha, y + \beta) * G_{(f_g, \theta_g)}(\alpha, \beta) \quad (1)$$

where f_g , θ_g and σ are the spatial frequency, orientation, and space constant of the Gabor envelope.

Texture features generated by Gabor filters have been increasingly considered and applied to document image analysis. During the last two decades, Gabor texture analysis approaches have been proposed for biometric identification based on handwriting [27, 28, 29], writer identification [30], handwritten word recognition [31], character recognition [32], font recognition [33], script identification [34, 35], signature recognition [36], palm print recognition [37], *etc.* Zhu *et al.* proposed a texture-analysis-based algorithm for automatic font recognition by extracting Gabor features [33]. The authors showed a 99,1% of mean recognition rate. Ma and Doermann proposed a Gabor filter based multi-class classifier in order to identify scripts, font-faces and font-styles [38].

Nevertheless, numerous approaches have been sought for text segmentation and extraction from digital document images using Gabor descriptors [39, 40, 41, 42]. Several studies have been conducted in the literature for page layout analysis using the multi-channel Gabor filters [43, 38, 44], while few ones have explored Gabor filters for ancient document image segmentation. For instance, Ribeiro *et al.* proposed an optical character recognition system for ancient document recognition by applying fuzzy methods on aligned oriented features extracted using Gabor filters in the training step [45]. Vieux and Domenger proposed a pixel-based classification approach to separate text from other classes (*e.g.* illustrations and background) by using a bank of Gabor filters at five scales (1, $\sqrt{2}$, 2, $2\sqrt{2}$ and 4) and six orientations ($k\frac{\pi}{6}, k \in \{0, \dots, 5\}$) [46]. They evaluated their approach on a public dataset containing magazines and technical journals. They found 86%, 82.7%, and 53.7% of F-measure for segmenting background, text, and graphic pixels, respectively. Jain *et al.* showed the effectiveness of applying a multi-channel Gabor filtering-based texture segmentation approach for the segmentation and classification of document images [47]. They chose the five following spatial frequencies: $4\sqrt{2}$, $8\sqrt{2}$, $16\sqrt{2}$, $32\sqrt{2}$ and $64\sqrt{2}$. Charrada and Ben Amara extracted nets from ancient Arab periodicals by exploring Gabor filters [48]. Zhong and Cheriet used the dimensionally reduced multi-channel Gabor filters for text block identification on image patches from ancient documents [49]. The authors extracted 28 Gabor filters from image patches in their experiments, where 7 spatial frequencies ($\sqrt{2}$, $2\sqrt{2}$, $4\sqrt{2}$, $8\sqrt{2}$, $16\sqrt{2}$, $32\sqrt{2}$ and $64\sqrt{2}$) and 4 orientation angles (0, $\pi/4$, $\pi/2$ and $3\pi/4$) were pre-defined. The four directions (0, $\pi/4$, $\pi/2$ and $3\pi/4$) are widely used in the literature [39, 47, 33, 38].

Designing the proper channels in order to generate filters tuned to several different frequencies and orientations has been illustrated as the crucial issue in using Gabor filters for texture characterization. Dunn *et al.* suggested an automatic approach for finding the optimal channels for discriminating distinct textures, but the computational complexity was very high [50, 51]. Bianconi and Fernández evaluated the impact of the Gabor filter parameters on texture classification [52]. They reported that an increase number of frequencies and orientations had an insignificant influence on texture classification performance. But, they confirmed that the best performance of texture classification is conditioned by the design of the convenient Gabor channels. Clausi and Jernigan presented a comparative study of different techniques used to extract Gabor descriptors for texture discrimination [53]. The authors showed that the magnitude response outperformed the other different evaluated methods, such as using only the real component, *etc.* Arivazhagan *et al.* introduced the rotation invariant features by computing the mean and variance of the Gabor filtered image for texture classification [54].

In this work, the magnitude response of the output of Gabor functions is investigated. The magnitude of the output is important if the specified Gabor filter matched the particular texture, otherwise low response to the specified Gabor filter corresponds to poor match of the dominant texture properties of the analyzed image to the set of the spatial-frequency components of the fixed Gabor filter [55]. By convoluting the analyzed whole document image at each specified channel defined by a pair of orientation and frequency, Gabor features are extracted. 24 Gabor filters are applied (6 distinct spatial frequencies $f_g = \{2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2} \text{ and } 64\sqrt{2}\}$ and 4 different orientations $\theta_g = \{0, \pi/4, \pi/2 \text{ and } 3\pi/4\}$). The space of Gabor filter is set constant $\sigma = \sigma_x = \sigma_y = 1$. Then, two simple statistics: the mean value and standard deviation of the Gabor filtered magnitude responses corresponding to all pixels defined in the analyzed sliding window of the filtered image, are extracted. They represent the statistical distribution of Gabor magnitude responses:

- **Mean:** This feature characterizes the average of the Gabor filtered magnitude responses corresponding to all pixels defined in the analyzed sliding window of the filtered image. This descriptor quantifies how match the dominant texture properties of the analyzed image to the set of spatial-frequency components of the fixed Gabor filter (*cf.* equation (2)).

- **Standard deviation:** This descriptor determines how much the dispersion from the computed mean of the Gabor filtered magnitude response exists (*cf.* equation (3)).

The two simple first order statistics representing Gabor features are detailed in Table 1.

Table 1. Gabor features.

Features	Expression
Mean	$F_{(f_g, \theta_g)}^{(1)} = \frac{\sum_{x=1}^M \sum_{y=1}^N I_{G_{(f_g, \theta_g)}}(x, y)}{M * N} \quad (2)$
Standard deviation	$F_{(f_g, \theta_g)}^{(2)} = \frac{\sum_{x=1}^M \sum_{y=1}^N [I_{G_{(f_g, \theta_g)}}(x, y) - F_{(f_g, \theta_g)}^{(1)}]^2}{M * N} \quad (3)$

In Table 1, $G_{e(f_g, \theta_g)}$ and $G_{o(f_g, \theta_g)}$ denote the spatial frequency responses of the even and odd symmetric Gabor filters.

$$G_{(f_g, \theta_g)}(\alpha, \beta) = \sqrt{[G_{e(f_g, \theta_g)}(\alpha, \beta)]^2 + [G_{o(f_g, \theta_g)}(\alpha, \beta)]^2}$$

$$G_{e(f_g, \theta_g)} = \frac{H_{1(f_g, \theta_g)}(\alpha, \beta) + H_{2(f_g, \theta_g)}(\alpha, \beta)}{2}$$

$$G_{o(f_g, \theta_g)} = \frac{H_{1(f_g, \theta_g)}(\alpha, \beta) - H_{2(f_g, \theta_g)}(\alpha, \beta)}{2j}$$

$$H_{1(f_g, \theta_g)}(\alpha, \beta) = \exp\{-2\pi\sigma^2[(\alpha - f_g \cos \theta_g)^2 + (\beta - f_g \sin \theta_g)^2]\}$$

$$H_{2(f_g, \theta_g)}(\alpha, \beta) = \exp\{-2\pi\sigma^2[(\alpha + f_g \cos \theta_g)^2 + (\beta - f_g \sin \theta_g)^2]\}$$

$$j^2 = -1$$

4. PROPOSED DOCUMENT ENHANCEMENT AND SEGMENTAION ALGORITHM

To ensure a relevant segmentation of graphical regions from textual ones on the one hand, and discrimination of text in a variety of situations of different fonts and scales on the other hand in ancient documents, a fast automatic ancient document enhancement and segmentation algorithm is proposed in this article. Figure 1 illustrates the detailed schematic block representation of the proposed algorithm.

Firstly, an ancient document image is fed as input (*cf.* Figure 2(a)) and is read as a gray-scale image. The different steps of the proposed algorithm is processed on gray-scale document images without introducing a binarizing task. A binarization step is avoided because it causes a loss of information specifically for information extraction and analysis. Then, to obtain enhanced backgrounds of noisy ancient documents and reduce the step complexity of a pixel-based segmentation of ancient documents, a foreground-background separation task based on SLIC segmentation [21] and k-means clustering [56] algorithms is performed.

By setting the number of superpixels k_s equal to 0.1% of image pixels, the SLIC technique is carried out on a gray-scale image and over-segmented image is produced. Afterwards, the background and foreground superpixels are classified based on computing the mean gray-level value of each superpixel, which is determined by averaging over all the gray-level pixels belonging to the superpixel region, and using the k-means algorithm. As a non-supervised clustering technique,

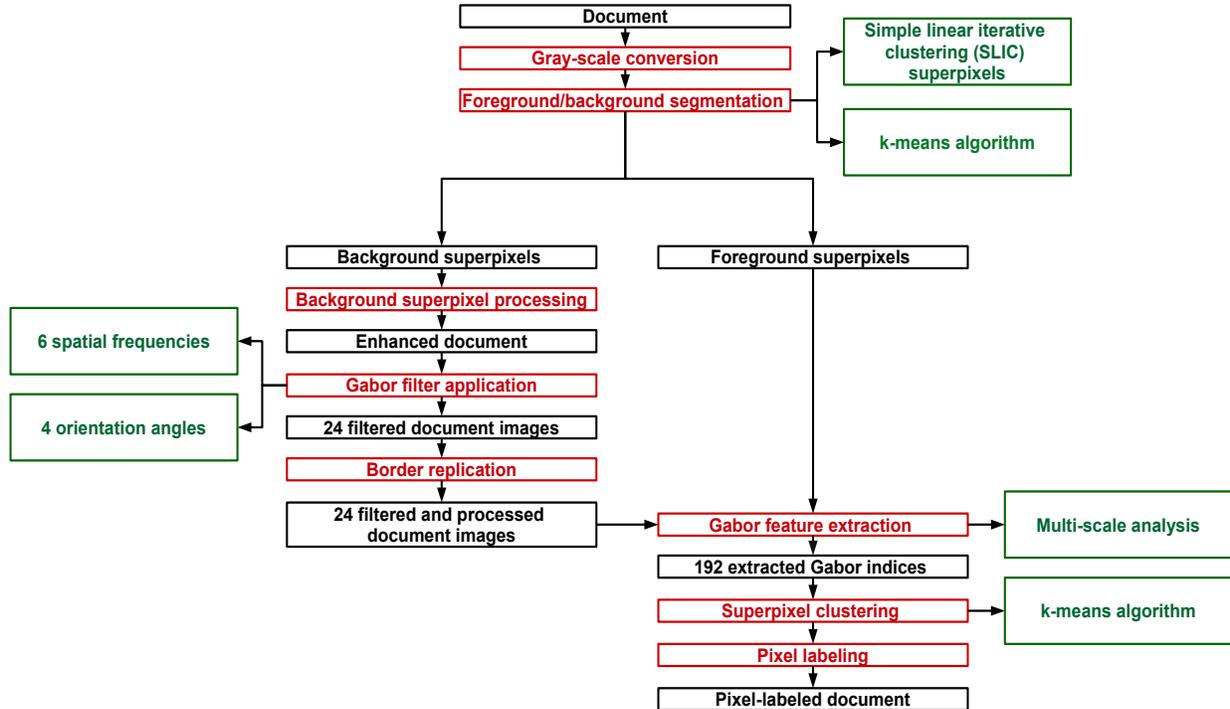


Figure 1. Proposed ancient document enhancement and segmentation algorithm.

the k-means algorithm is performed on the computed mean gray-level values of superpixels, without taking into account the image spatial coordinates, by setting the number of clusters k_c equal to 2 to extract two clusters and subsequently to segment an image into two layers. One represents the information of the background (*cf.* Figure 2(c)) and the other represents the foreground (*e.g.* noise, text fields, drawings, *etc.*) (*cf.* Figure 2(d)). Since the k-means algorithm is carried out, the background superpixels of the original gray-level image are only processed by assigning the value of a white pixel (*i.e.* a 255 gray-level value) to their centers and the pixels belonging to them. However, the values of the gray-level foreground superpixels and their pixels of the original gray-level image are remained unchangeable. Thus, an enhanced and non-noisy background is achieved (*cf.* Figure 2(e)). Figure 2(e) illustrates an example of enhanced image by the superpixel technique with a clean background.

Afterwards, 24 Gabor filters, which are obtained by using 4 different orientations $\theta_g = \{0, \pi/4, \pi/2 \text{ and } 3\pi/4\}$ and 6 distinct spatial frequencies $f_g = \{2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2} \text{ and } 64\sqrt{2}\}$, are applied to the enhanced gray-level ancient document image. When convolving the enhanced gray-level ancient document image with 24 Gabor filters, 24 magnitude responses are generated (*i.e.* 24 transformed images by the selective Gabor filters are produced). Then, a quick and easy way to compute Gabor features on the whole transformed image by the selective Gabor filter, is to introduce a border replication step before the texture extraction task. By using rectangular overlapping processing windows, Gabor descriptors are only extracted from the selected foreground superpixels of the transformed image by the selective Gabor filter and the border replication step, at four different sizes of sliding windows: (16×16) , (32×32) , (64×64) and (128×128) to adopt a multi-scale approach. Thus, a feature vector (with dimension 48 to represent 24 Gabor filters) is produced based on the computed mean and standard deviation of the magnitude response of the transformed image by the selective Gabor filter which are extracted from one analyzed sliding window. Thus, a total of 48 features are extracted from each foreground superpixel defined in one analyzed sliding window. A 192-dimensional feature vector is subsequently formed through four different sizes of sliding windows.

Then, a foreground superpixel clustering task is performed by partitioning Gabor-based feature sets into compact and well-separated clusters in the feature space for segmenting graphical regions from textual ones on the one hand, and discriminating text in a variety of situations of different fonts and scales on the other hand. The foreground superpixel clustering task does not include spatial information and is performed by using the k-means algorithm.

Finally, a phase of labeling clusters of the gray-level foreground superpixels and the gray-level pixels belonging to each superpixel in the enhanced document image is carried out with respect to the results of the superpixel clustering phase.

Since the clustering and labeling phases of the proposed algorithm had been performed, the homogeneous regions in the ancient document are determined by labeling clusters or groups of pixels which share similar textural characteristics (cf. Figure 2(f)).



Figure 2. Illustration of the intermediate results of the different steps of the proposed ancient document enhancement and segmentation algorithm: figure (a) shows an example of an ancient document image (as an input of the proposed algorithm). Figure (b) illustrates its corresponding segmented image (by using the SLIC superpixel technique) and a zoomed region of it. Figures (c) and (d) show the background and foreground SLIC superpixels, respectively. Colors assigned to the background (foreground respectively) superpixels which are illustrated in Figure (c) ((d) respectively) are randomly generated. Figure (e) depicts an enhanced image and a zoomed region of it. Figure (f) illustrates the pixel-labeled image and a zoomed region of it (as an output of the proposed algorithm (graphic regions (blue), textual regions (green))).

5. EXPERIMENTS

We have experimentally evaluated the proposed ancient document enhancement and segmentation algorithm on 500 pages of ancient documents. In this section, we discuss the performance of the proposed algorithm in detail after describing our experimental corpus and its associated ground-truth.

5.1 Corpus and preparation of ground-truth

Our corpus is structured into two categories of real scanned historical documents differentiated by their content (*cf.* Figure 3), reflecting the challenges of our work to determine if texture features can segment efficiently graphical regions from textual ones on the one hand, and discriminate text in a variety of situations of different fonts and scales on the other hand. Our experimental corpus includes a sufficient number of historical documents with both simple and complex layouts in each of the content categories which are ground-truthed to ensure a better understanding of the behavior of the extracted Gabor features. It is composed of:

- 250 pages containing graphics and one text font (*cf.* Figure 3(a))
- 250 pages containing only two fonts (*cf.* Figure 3(b))

The ground-truth for document images are manually outlined using rectangular regions drawn around each selected zone. The zones are ground-truthed by zoning manually each content type, *i.e.* each rectangular region is classified into text or graphics. Different labels for regions with different fonts are also defined for evaluating the performance of Gabor feature to separate distinct text fonts. Ground-truth is performed using the ground-truthing editor, Ground-truthing Environment for Document Images (GEDI), a public domain document image annotation tool that labels spatial boundaries of regions* [57]. By specifying rectangular regions on a document image and assigning them to one of the many pre-defined content types, GEDI generates an XML schema representing the location on the page, height, width and label of each region (*cf.* Figure 3(c)).

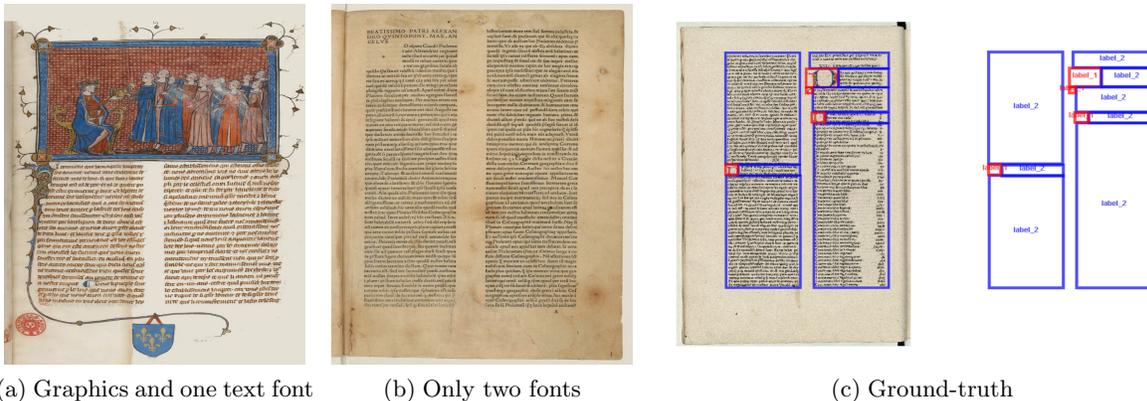
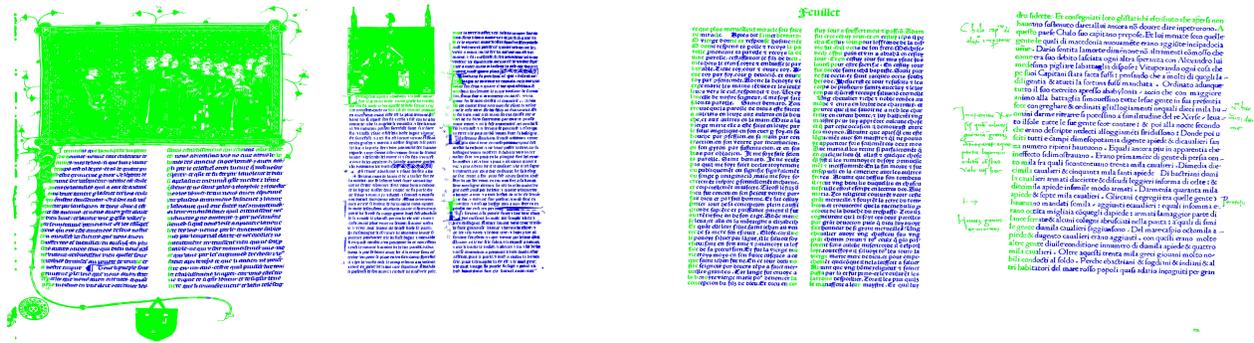


Figure 3. Illustration of some examples of our experimental corpus and its associated ground-truth. Figure (a) illustrates an example of an ancient document containing graphics and one text font. Figure (b) shows an example of an ancient document containing only two text fonts. Figure (c) depicts an example of an associated ground-truth to an ancient printed document containing graphics and one text font.

5.2 Evaluation and results

The results of the proposed ancient document enhancement and segmentation algorithm using the SLIC superpixel, multi-channel Gabor filter and k-means techniques are illustrated in Figures 4(a) and 4(b). By visual inspection of the pixel-labeled images, we note that better results are obtained with ancient documents containing graphics and one text font (*cf.* Figure 4(a)) than those containing only two text fonts (*cf.* Figure 4(b)). The proposed algorithm provides very satisfying results particularly in distinguishing textual regions from graphical ones. This highlights a much greater discriminant power for separating text and graphic regions (*cf.* Figure 4(a)) than for distinguishing two text fonts (*cf.* Figure 4(b)). Nevertheless, the extracted Gabor features confuse the manuscript notes and printed text on the one hand, and two distinct text font with different sizes on the other hand (*cf.* Figure 4(b)). This confusion can be explained by the limitations of Gabor descriptors to segregate close distinct kinds of information (*i.e.* the horizontal and vertical spacings between two text fonts are small) since Gabor indices are extracted for a specified range of frequency and direction values.

*<http://gedigroundtruth.sourceforge.net/>



(a) Graphics and one text font

(b) Only two fonts

Figure 4. Examples of results of the proposed document enhancement and segmentation algorithm. The pixel labeling of the extracted Gabor features is performed from enhanced ancient documents using the k-means algorithm by setting the number of homogeneous and similar content regions (*i.e.* the number of clusters) to the one defined in our ground-truth, which is equal to 2. Figures (a) and (b) depict four result examples of the case when Gabor features are extracted and analyzed from the selected foreground SLIC superpixels. Figure (a) illustrates some examples of ancient document containing graphics and one text font (graphic regions (green), textual regions (blue)) while figure (b) shows some examples of ancient document containing only two fonts. Since the process is unsupervised, the colors attributed to text or graphics may differ from one document image to another.

This way of assessing the effectiveness of a segmentation method is inherently a subjective evaluation and we need to assess and evaluate robustness using an appropriate quantitative metric. Thus, in this study F-measure (F) is computed. To get an insight into the classification accuracy, a confusion matrix (error matrix or contingency table) [58] is usually computed. From the confusion matrix, F is deduced [59]. This accuracy metric is related to how representative the clusters are of classes and help to determine classes which are not able to segregate groups of data and give an insight into the confusion and mis-classification rates.

The computed accuracy classification values are congruent and very promising. 77% of mean F-measure is obtained by evaluating 500 pages of ancient documents. 82% and 72% of mean F-measure are noted for documents containing graphics and one text font, and only two fonts, respectively. Promising accuracy results are obtained with pages containing graphics and one text font, compared to pages containing only two fonts. So this confirmed our hypothesis that Gabor attributes have a much greater discriminant power for separating single text font and graphic regions than for separating two distinct text fonts. Nevertheless, the classification result could be improved by integrating a processing stage after the pixel labeling task, which consists of pixel grouping by taking into consideration the topographical relationships of pixels and their labels.

Table 2. Evaluation of the proposed ancient document enhancement and segmentation algorithm by computing F-measure (F). $\mu(F)$ and $\sigma(F)$ represent the mean and standard deviation values of F , respectively when extracting and analyzing Gabor features. The “ μ^* ” value is obtained by extracting and analyzing Gabor features from the selected foreground SLIC superpixels. The “ μ^{**} ” value is obtained by extracting and analyzing Gabor features from all foreground pixels.

	$\mu^*(F)$	$\sigma^*(F)$	$\mu^{**}(F)$	$\sigma^{**}(F)$
One font and graphics	0.82	0.15	0.64	0.15
Only two fonts	0.72	0.12	0.62	0.12
Overall	0.77	0.13	0.63	0.13

The experiment was run on a SGI Altix ICE 8200 cluster (1 CPU and 2 gigabytes (GB) allocated memory on a Quad-Core X5355@2.66GHz running Linux). The time required to process a page (1965 × 2750 pixels) using the proposed algorithm with extracting and analyzing Gabor features from the selected foreground SLIC superpixels, and from all foreground pixels is 04' 40" and 32' 36", respectively. The amount of memory consumed when using the proposed algorithm with extracting and analyzing Gabor features from all foreground pixels is 5 times higher than with extracting and analyzing Gabor features from the selected foreground SLIC superpixels. This confirmed that the use of

SLIC superpixels for analyzing Gabor features is an interesting alternative since the data cardinality is reduced and a significant gain in computation time and memory is obtained.

A visually comparison of the results (*cf.* Figure 5) when extracting and analyzing Gabor features from all foreground pixels (*cf.* Figure 5(b)) and from the selected foreground SLIC superpixels (*cf.* Figures 5(a)), indicates that better results are obtained with the case of extracting and analyzing Gabor features from the selected foreground SLIC superpixels. The outperformace obtained when extracting and analyzing Gabor features from the selected foreground SLIC superpixels highlights our choice of using the SLIC superpixels for both document enhancement and segmentation (*i.e.* Gabor feature analysis). In conclusion, we observe a drop in the average of 14%(F) when extracting and analyzing Gabor features from all foreground pixels. We have demonstrated both qualitatively and quantitatively the effectiveness of using the SLIC superpixels for both document enhancement and segmentation.

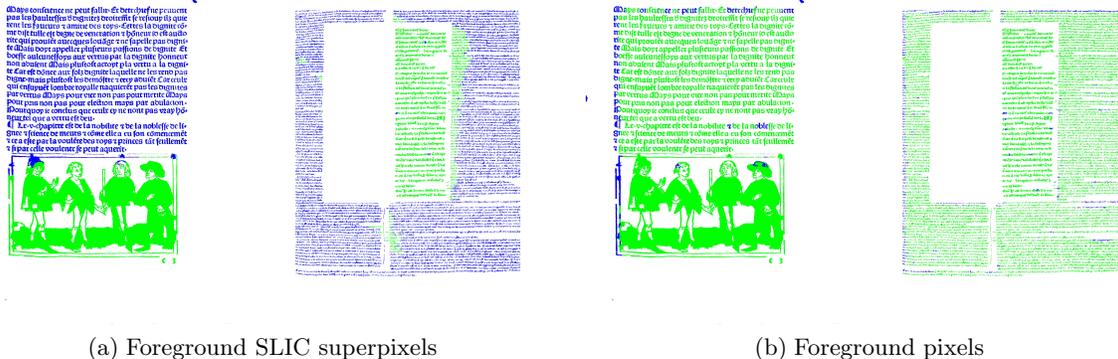


Figure 5. Examples of results of the pixel labeling of Gabor features when they are extracted and analyzed from the selected foreground SLIC superpixels and all foreground pixels. Figures (a) and (b) depict four result examples when Gabor features are extracted and analyzed from the selected foreground SLIC superpixels and from all foreground pixels, respectively. The pixel labeling is performed from enhanced ancient documents using the k-means algorithm by setting the number of homogeneous and similar content regions (*i.e.* the number of clusters) to the one defined in our ground-truth, which is equal to 2.

6. CONCLUSIONS AND FUTURE WORK

This article proposes an ancient document enhancement and segmentation algorithm with no hypothesis concerning the document model or the typographical parameters of the document structure. The goal of this algorithm is to find homogeneous regions within the content of digitized historical document images by extracting and analyzing texture features independently of the layout of the pages. This algorithm is based on SLIC superpixel, multi-channel Gabor filter and k-means techniques. It was evaluated on 500 pages of ancient documents with promising results. We conclude that Gabor features provide a good discrimination of the foreground layers of a document, particularly between text and graphics.

The first aspect of future work will be to use the proposed algorithm on a larger database and a wider variety of ancient documents (*i.e.* documents containing more than two distinct text fonts and various graphic types). This study is ongoing and will evaluate the algorithm more adequately, with more convincing experimental results. We will then study and combine statistical, geometric, model-based and other frequency texture features in order to refine the segmentation and ensure a distinction between different text fonts and various graphic types. Moreover, our results could be improved if we included fuzzy clustering methods in our clustering approach. In addition, a clustering methodology will be introduced to estimate automatically the correct number of clusters (*e.g.* analysis of changes in average silhouette width values computed from clusters built by using the k-means algorithm). Furthermore, by integrating a new processing stage after pixel labeling, which consists of pixel grouping that takes into consideration the topographical relationships of pixels and their labels (*e.g.* some operators from mathematical morphology) the pixel labeling results should be improved. This will be studied in the near future.

ACKNOWLEDGMENTS

This work is supported by the ANR (French National Research Agency), under Grant ANR-10-CORD-0020, which is gratefully acknowledged. The authors would like also to thank Geneviève Cron for providing access to the Gallica digital library.

REFERENCES

- [1] Baird, H. S., “Digital libraries and document image analysis,” in [*International Conference on Document Analysis and Recognition*], 2–14, IEEE (2003).
- [2] Coustaty, M., Raveaux, R., and Ogier, J. M., “Historical document analysis: A review of French projects and open issues,” in [*European Signal Processing Conference*], 1445–1449, EURASIP (2011).
- [3] Okun, O. and Pietikäinen, M., “A survey of texture-based methods for document layout analysis,” in [*Workshop on Texture Analysis in Machine Vision*], 137–148, Springer-Verlag (1999).
- [4] Uttama, S., Loonis, P., Delalandre, M., and Ogier, J. M., “Segmentation and retrieval of ancient graphic documents,” in [*International Workshop on Graphics Recognition on Graphics Recognition (GREC): Ten Years Review and Future Perspectives*], 88–98, Springer-Verlag (2006).
- [5] Diem, M. and Sablatnig, R., “Recognition of degraded handwritten characters using local features,” in [*International Conference on Document Analysis and Recognition*], 221–225, IEEE (2009).
- [6] Likforman-Sulem, L., Zahour, A., and Taconet, B., “Text line segmentation of historical documents: A survey,” *International Journal of Document Analysis and Recognition* , 123–138 (2007).
- [7] Fischer, A., Baechler, M., Garz, A., Liwicki, M., and Ingold, R., “A combined system for text line extraction and handwriting recognition in historical documents,” in [*International Workshop on Document Analysis Systems*], 71–75, IEEE (2014).
- [8] Baird, H. S. and Moll, M. A., “Document content inventory and retrieval,” in [*International Conference on Document Analysis and Recognition*], 93–97, IEEE (2007).
- [9] An, C., Baird, H. S., and Xiu, P., “Iterated document content classification,” in [*International Conference on Document Analysis and Recognition*], 252–256, IEEE (2007).
- [10] An, C., Yin, D., and Baird, H. S., “Document segmentation using pixel-accurate ground truth,” in [*International Conference on Pattern Recognition*], 245–248, IEEE (2010).
- [11] Kise, K., [*Page segmentation techniques in document analysis*], Handbook of Document Image Processing and Recognition, Springer-Verlag (2014).
- [12] Petrou, M. and Sevilla, P. G., [*Image processing: Dealing with texture*], John Wiley & Sons (2006).
- [13] Haralick, R. M., Shanmugam, K., and Dinstein, I., “Textural features for image classification,” *Systems Man and Cybernetics* , 610–621 (1973).
- [14] Gabor, D., “Theory of communication. Part 1: The analysis of information,” *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering* , 429–441 (1946).
- [15] Mehri, M., Gomez-Krämer, P., Héroux, P., Boucher, A., and Mullot, R., “Texture feature evaluation for segmentation of historical document images,” in [*International Workshop on Historical Document Imaging and Processing*], 102–109, ACM (2013).
- [16] Agam, G., Bal, G., Frieder, G., and Frieder, O., “Degraded document image enhancement,” in [*Document Recognition and Retrieval*], SPIE (2007).
- [17] Wu, C. H. and Chang, H. H., “Gaussian noise estimation with superpixel classification in digital images,” in [*International Congress on Image and Signal Processing*], 373–377, IEEE (2012).
- [18] Fulkerson, B., Vedaldi, A., and Soatto, S., “Class segmentation and object localization with superpixel neighborhoods,” in [*International Conference on Computer Vision*], 670–677, IEEE (2009).
- [19] Achanta, R., Shaji, A., Lucchi, A., Fua, P., and Süsstrunk, S., “SLIC superpixels compared to state-of-the-art superpixel methods,” *Pattern Analysis and Machine Intelligence* , 2274–2282 (2012).
- [20] Cohen, R., Asi, A., Kedem, K., El-Sana, J., and Dinstein, I., “Robust text and drawing segmentation algorithm for historical documents,” in [*International Workshop on Historical Document Imaging and Processing*], 110–117, ACM (2013).
- [21] Liu, M. Y., Tuzel, O., Ramalingam, S., and Chellappa, R., “Entropy rate superpixel segmentation,” in [*Computer Vision and Pattern Recognition*], 2097–2104, IEEE (2011).
- [22] Mehri, M., Kieu, V. C., Mhiri, M., Héroux, P., Gomez-Krämer, P., Mahjoub, M. A., and Mullot, R., “Robustness assessment of texture features for the segmentation of ancient documents,” in [*International Workshop on Document Analysis Systems*], 293–297, IEEE (2014).
- [23] Campbell, F. W. and Robson, J. G., “Application of Fourier analysis to the visibility of gratings,” *The Journal of Physiology* , 551–566 (1968).
- [24] Daugman, J. G., “Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters,” *Journal of the Optical Society of America A* , 1160–1169 (1985).
- [25] Ursani, A. A., Kpalma, K., and Rosin, J., “Texture features based on Fourier transform and Gabor filters: An empirical comparison,” in [*International Conference on Machine Vision*], 67–72, IEEE (2007).
- [26] Jain, A. K. and Farrokhnia, F., “Unsupervised texture segmentation using Gabor filters,” *Pattern Recognition* , 1167–1186 (1991).
- [27] Zhu, Y., Tan, T., and Wang, Y., “Biometric personal identification based on handwriting,” in [*International Conference on Pattern Recognition*], 797–800, IEEE (2000).

- [28] Feddaoui, N. and Hamrouni, K., "Personal identification based on texture analysis of Arabic handwriting text," in [*International Conference on Information & Communication Technologies*], 1302–1307, IEEE (2006).
- [29] Chen, J., Cao, H., Prasad, R., Bhardwaj, A., and Natarajan, P., "Gabor features for offline Arabic handwriting recognition," in [*International Workshop on Document Analysis Systems*], 53–58, ACM (2010).
- [30] Bensefia, A., Paquet, T., and Heutte, L., "A writer identification and verification system," *Pattern Recognition Letters*, 2080–2092 (2005).
- [31] Buse, R., Liu, Z. Q., and Caelli, T., "A structural and relational approach to handwritten word recognition," *Systems Man and Cybernetics*, 847–861 (1997).
- [32] Wang, X., Ding, X., and Liu, C., "Gabor filters-based feature extraction for character recognition," *Pattern Recognition*, 369–379 (2005).
- [33] Zhu, Y., Tan, T., and Wang, Y., "Font recognition based on global texture analysis," *Pattern Analysis and Machine Intelligence*, 1192–1200 (2001).
- [34] Tan, T. N., "Rotation invariant texture features and their use in automatic script identification," *Pattern Analysis and Machine Intelligence*, 751–756 (1998).
- [35] Joshi, G. D., Garg, S., and Sivaswamy, J., "Script identification from Indian documents," in [*International Workshop on Document Analysis Systems*], 255–267, Springer-Verlag (2006).
- [36] Kekre, H. B. and Bharadi, V. A., "Gabor filter based feature vector for dynamic signature recognition," *International Journal of Computer Applications*, 74–80 (2010).
- [37] Mu, M. and Ruan, Q., "Mean and standard deviation as features for palmprint recognition based on Gabor filters," *International Journal of Pattern Recognition and Artificial Intelligence*, 491–512 (2011).
- [38] Ma, H. and Doermann, D., "Gabor filter based multi-class classifier for scanned document images," in [*International Conference on Document Analysis and Recognition*], 968–972, IEEE (2003).
- [39] Jain, A. K. and Bhattacharjee, S., "Text segmentation using Gabor filters for automatic document processing," *Machine Vision and Applications*, 169–184 (1992).
- [40] Chan, W. and Coghill, G., "Text analysis using local energy," *Pattern Recognition*, 2523–2532 (2001).
- [41] Raju, S. S., Pati, P. B., and Ramakrishnan, A. G., "Gabor filters for document analysis in Indian bilingual documents," in [*International Workshop on Document Image Analysis for Libraries*], 233–243, IEEE (2004).
- [42] Raju, S., Pati, P., and Ramakrishnan, A., "Text localization and extraction from complex color images," in [*International Symposium on Visual Computing*], 486–493, Springer-Verlag (2005).
- [43] Randen, T. and Husøy, J. H., "Segmentation of text/image documents using texture approaches," (1994).
- [44] Pati, P. B., Raju, S. S., Pati, N., and Ramakrishnan, A. G., "Gabor filters for document analysis in Indian bilingual documents," in [*International Conference on Intelligent Sensing and Information Processing*], 123–126, IEEE (2004).
- [45] Ribeiro, C. S., Gil, J. M., Pinto, J. R. C., and da Costa Sousa, J. M., "Ancient document recognition using fuzzy methods," in [*International Conference on Fuzzy Systems*], 833–838, IEEE (2005).
- [46] Vieux, R. and Domenger, J. P., "Hierarchical clustering model for pixel-based classification of document images," in [*International Conference on Pattern Recognition*], 290–293, IEEE (2012).
- [47] Jain, A. K. and Zhong, Y., "Page segmentation using texture analysis," *Pattern Recognition*, 743–770 (1996).
- [48] Charrada, M. A. and Amara, N. E. B., "Texture approach for nets extraction application to old Arab newspapers images structuring," in [*Image Processing Theory, Tools and Applications*], 212–216, IEEE (2012).
- [49] Zhong, G. and Cheriet, M., "Image patches analysis for text block identification," in [*International Conference on Information Science, Signal Processing and their Applications*], 1241–1246, IEEE (2012).
- [50] Dunn, D., Higgins, W. E., and Wakeley, J., "Texture segmentation using 2-D Gabor elementary functions," *Pattern Analysis and Machine Intelligence*, 130–149 (1994).
- [51] Dunn, D. and Higgins, W. E., "Optimal Gabor filters for texture segmentation," *Image Processing*, 947–964 (1995).
- [52] Bianconi, F. and Fernández, A., "Evaluation of the effects of Gabor filter parameters on texture classification," *Pattern Recognition*, 3325–3335 (2007).
- [53] Clausi, D. A. and Jernigan, M. E., "Designing Gabor filters for optimal texture separability," *Pattern Recognition*, 1835–1849 (2000).
- [54] Arivazhagan, S., Ganesanb, L., and Priyal, S. P., "Texture classification using Gabor wavelets based rotation invariant features," *Pattern Recognition Letters*, 1976–1982 (2006).
- [55] Bovik, A. C., Clark, M., and Geisler, W. S., "Multichannel texture analysis using localized spatial filters," *Pattern Analysis and Machine Intelligence*, 55–73 (1990).
- [56] MacQueen, J. B., "Some methods for classification and analysis of multivariate observations," in [*Berkeley Symposium on Mathematical Statistics and Probability*], 281–297, University of California Press (1967).
- [57] Doermann, D., Zotkina, E., and Li, H., "GEDII - A Groundtruthing Environment for Document Images," in [*International Workshop on Document Analysis Systems*], ACM (2010).
- [58] Jensen, J. R., [*Introductory digital image processing*], Prentice-Hall, Englewood Cliffs, NJ (1986).
- [59] Makhoul, J., Kubala, F., Schwartz, R., and Weischedel, R., "Performance measures for information extraction," in [*DARPA Broadcast News Workshop*], 249–252, Morgan Kaufmann Publishers, Inc (1999).