Learning Texture Features for Enhancement and Segmentation of Historical Document Images

Maroua Mehri, Nibal Nayef, Pierre Héroux, Petra Gomez-Krämer, Rémy Mullot

Learning Texture Features for Enhancement and Segmentation of Historical Document Images

Maroua Mehri
L3i, University of La Rochelle
La Rochelle, France
maroua.mehri@univ-lr.fr

Nibal Nayef
L3i, University of La Rochelle
La Rochelle, France
nibal.nayef@univ-lr.fr

Pierre Héroux
LITIS, University of Rouen
Saint-Etienne-du-Rouvray, France
pierre.heroux@univ-rouen.fr

Petra Gomez-Krämer
L3i, University of La Rochelle
La Rochelle, France
petra.gomez@univ-lr.fr

Rémy Mullot
L3i, University of La Rochelle
La Rochelle, France
remy.mullot@univ-lr.fr

ABSTRACT

Many challenges and open issues related to the tremendous growth in digitizing collections of cultural heritage documents have been raised, such as information retrieval in digital libraries or analyzing page content of historical books. Recently, graphic/text segmentation in historical documents has posed specific challenges due to many particularities of historical document images (e.g. noise and degradation, presence of handwriting, overlapping layouts, great variability of page layout). To cope with those challenges, a method based on learning texture features for historical document image enhancement and segmentation is proposed in this article. The proposed method is based on using the simple linear iterative clustering (SLIC) superpixels, Gabor descriptors and support vector machines (SVM). It has been evaluated on 100 document images which have been selected from the databases of the competitions (*i.e.* historical document layout analysis and historical book recognition) in the context of ICDAR conference and HIP workshop (2011 and 2013). To demonstrate the enhancement and segmentation quality, the evaluation is based on manually labeled ground truth and shows the effectiveness of the proposed method through qualitative and numerical experiments. The proposed method provides interesting results on historical document images having various page layouts and different typographical and graphical properties.

Categories and Subject Descriptors

I.4.6 [**Segmentation**]: Pixel classification; I.7.5 [**Document Capture**]: Document analysis

Keywords

Historical document images, enhancement, segmentation, SLIC superpixels, learning texture features, multi-scale technique.

1. INTRODUCTION

Document image analysis (DIA) has been a thriving topic of major interest of many researchers and one of the most explored fields in image analysis [25]. It consists of dividing a document image (DI) layout according to the nature of the extracted structure such as separating text from non-text regions or partitioning text into columns, text blocks, lines, words, *etc.* It starts by segmenting a DI in order to find and classify homogeneous regions or zones, such as graphical and textual regions [26]. Finding graphical regions can be used to segment and analyze the graphical part of historical heritage such as the drop caps [27], while determining text zones can be used as a pre-processing stage for character recognition [10], text line extraction [20], handwriting recognition [12], *etc.*

Several scientific works in contemporary DIA have described several relevant methods enabling multiple forms of indexing based on content analysis of DIs. Nevertheless, the transposition of these methods for historical DIA, that are dedicated initially for contemporary DIA, is not straightforward. Grana *et al.* [14] stated that, despite that state of the art methods have yielded reliable results for contemporary DIA, analyzing historical document images (HDIs) by separating textual regions from the graphical ones is still more challenging due to many particularities of HDIs (e.g. large variability of page layout, noise and degradation, page skew, complicated layout, random alignment, specific fonts, presence of embellishments, variations in spacing between the characters, words, lines, paragraphs and margins, overlapping object boundaries, superimposition of information layers). Indeed, processing HDIs usually includes several stages: preprocessing, analysis, segmentation and characterization [19]. For the problem of historical DIA, the main challenge is to analyze HDIs and to characterize their layouts and contents under significant degradation levels and different noise types and with no *a priori* knowledge about the layout, content, typography, font styles, scanning resolution or DI size, *etc.*

Antonacopoulos *et al.* [4] pointed out the significant need for robust and accurate DIA methods that deal with the idiosyncrasies of HDIs. In addition, Crasson and Fekete [9] highlighted the real need for automatic processing of digitized HDIs (HDI layout analysis and text/non-text separation) to facilitate the analysis and navigation in the corpus

of ancient manuscripts. Moreover, Kise [17] stated that the analysis of pages with constrained layouts (e.g. rectangular, Manhattan) and clean DIs has almost been solved while historical DIA is still an open problem due to the HDI particularities. Recently, few layout analysis methods for historical documents have been proposed in the literature in order to characterize the DI layout. For instance, Garz *et al.* [13] proposed a part-based detection of layout entities in ancient manuscripts using a multi-stage algorithm based on interest points. Nevertheless, Kise [17] 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. Hence, texture-based methods address the challenges of the existing state of the art ones. The use of texture-based methods for DIA has been shown to be effective with skewed and degraded images [23]. Given that there are significant degradations and no hypothesis concerning the layout, the graphical properties or typographical parameters of the analyzed HDI, the use of texture analysis techniques for HDI has become an appropriate choice. Recently, the superpixel approach has gained great attention of many researchers in document image analysis fields. For instance, Cohen *et al.* [8] separated drawings from background and noise of ancient documents by using spatial and color features which were extracted from superpixels. Asi *et al.* [7] proposed a learning-free approach to detect the main text area from side-notes in ancient manuscripts based on a coarse-to-fine scheme. First, a coarse segmentation of the main text area was processed by using Gabor filters (GFs). Then, the segmentation was refined by formulating the problem as an energy minimization task and achieving the minimum using graph cuts. Wei *et al.* [28] compared three classifiers based on support vector machines (SVM), multi-layer perceptron (MLP) and Gaussian mixture models (GMM) to detect physical structure of HDIs. They concluded that both SVM and MLP classifiers had better performance than GMM. Pixels were classified into four classes: periphery, background, text or decoration, in the first classification level. Then, the three evaluated classifiers were combined together to ensure a vote for the pixel label in order to further improve the pixel-labeling results.

In this article, a method based on learning texture features for HDI enhancement and segmentation is proposed to assist the analysis of HDIs. The remainder of this article is organized as follows. The proposed enhancement and segmentation algorithm for HDIs based on the use of the simple linear iterative clustering (SLIC) superpixels, Gabor descriptors and SVM is detailed in Section 2. Section 3 describes firstly the experimental protocol by presenting the corpus and the defined ground truth (*cf.* Section 3.1). Secondly, to evaluate the performance of the proposed algorithm, a set of experiments on the *"HBR2013 dataset"* which have been provided by the "Centre of competence in digitisation"[1] IMPACT research team in the context of ICDAR conference and HIP workshop (2011 and 2013) is detailed in Section 3.2. Qualitative and numerical results are given to demonstrate the enhancement and segmentation quality. Our discussion, conclusions and future work are presented in Sections 4 and 5.

[1] `http://digitisation.eu`

2. THE PROPOSED METHOD

In this work, we are not looking for an accurate segmentation, but to find regions with similar textural content as easily, quickly and automatically as possible based on the pixel labeling results. It has been largely proved that the texture-based analysis methods are relevant for DIA and characterization [16, 22]. But, it can neither segment a DI into graphics, paragraphs, *etc.* nor characterize its structure (e.g. columns, rows, paragraphs). The region segmentation and classification tasks can be carried at the end after introducing a post-processing phase by taking into consideration the topological or spatial relationships (e.g. hierarchy, inclusion, neighborhood position). The proposed method based on learning texture features for HDI enhancement and segmentation has the possibility to be extended for consequent DI processing such as region segmentation and classification, by introducing a standard post-processing method (e.g. morphological cleaning approach, multi-scale majority voting technique). Figure 1 illustrates the detailed schematic block representation of the proposed method based on learning texture features for HDI enhancement and segmentation.

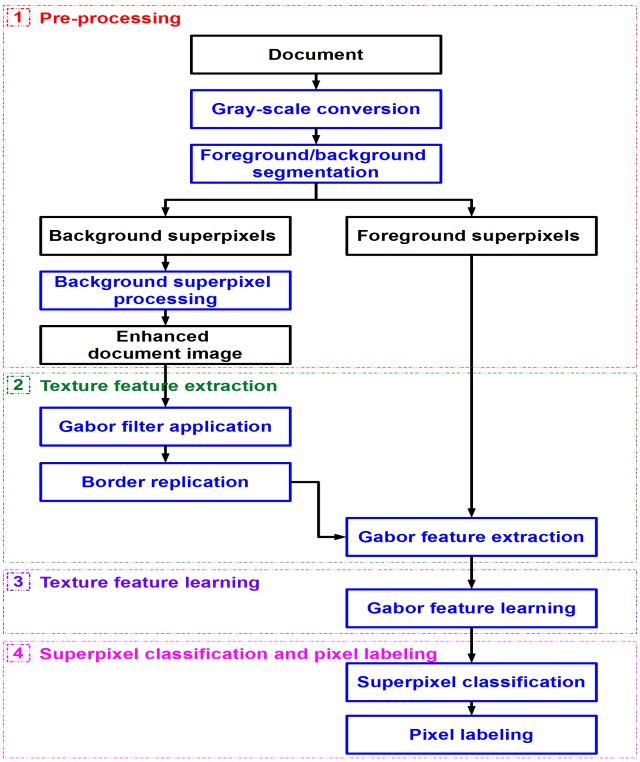


Figure 1: Flowchart of the proposed method based on learning texture features for HDI enhancement and segmentation.

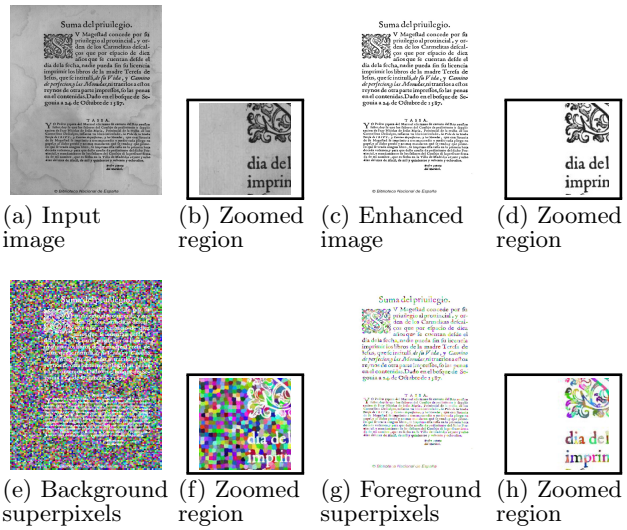
The proposed texture learning method for HDI enhancement and segmentation is conceptualized by four modular processes:

1. ***Pre-processing*** (*cf.* Section 2.1),
2. ***Texture feature extraction*** (*cf.* Section 2.2),
3. ***Texture feature learning*** (*cf.* Section 2.3),

4. ***Superpixel classification and pixel labeling*** (*cf.* Section 2.4).

2.1 Pre-processing

Firstly, a HDI is fed as input (*cf.* Figure 2(a)) and is read as a gray-scale image.



(a) Input image (b) Zoomed region (c) Enhanced image (d) Zoomed region

(e) Background superpixels (f) Zoomed region (g) Foreground superpixels (h) Zoomed region

Figure 2: Illustration of the intermediate results of the different steps of the pre-processing task. Figures (a) and (b) show an example of a HDI (as an input of the proposed algorithm) and a zoomed region of it, respectively. Figures (c) and (d) illustrate the enhanced HDI (*i.e.* the resulting image of the enhancement step of the proposed algorithm) and a zoomed region of it, respectively. Figures (e) and (f) depict the background SLIC superpixels and a zoomed region of it, respectively. Figures (g) and (h) illustrate the foreground SLIC superpixels and a zoomed region of it, respectively. Colors assigned to the background (foreground respectively) superpixels which are illustrated in Figure (e) ((g) respectively) are randomly generated.

Secondly, a foreground/background superpixel-based segmentation is carried out by means of the SLIC superpixel technique. Instead of using a rigid structure of pixel grid for pixel-based feature extraction and analysis, the superpixel technique has been used in our method as the basic unit when extracting texture features. The superpixel technique 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 [1]. By using the SLIC superpixel technique in the proposed method, pixels sharing similar characteristics or properties (e.g. texture cues, contour, color) are grouped into a significant polygon-shaped region. Indeed, by setting the number of SLIC superpixels k_s equal to 0.002% of image pixels, an over-segmented image representing a compact content map is generated. Afterwards, to segment an image into two layers (*i.e.*, foreground and background), the background and foreground superpixels are classified using the k-means algorithm based on computing the mean gray-level value of each superpixel. The mean gray-level value of each superpixel is determined by averaging over all the gray-level pixels belonging to the superpixel region. The k-means algorithm is performed on the computed mean gray-level values of SLIC superpixels, without taking into account the image spatial coordinates and by setting the number of clusters k_c equal to 2 to extract two clusters. One represents the information of the background (*cf.* Figure 2(e)) and the other represents the foreground (e.g. noise, text fields, drawings) (*cf.* Figure 2(g)). Since the foreground/background superpixel-based segmentation step 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.* 255 gray-level value) to their centers and the pixels belonging to them. Nevertheless, the values of the gray-level foreground superpixels and their pixels of the original gray-level image remain unchanged. Thus, an enhanced and non-noisy background is obtained (*cf.* Figure 2(c)). Figure 2(c) depicts an example of an enhanced image by the superpixel technique with a clean background.

2.2 Texture Feature Extraction

Six well-known and widely used texture-based feature sets (autocorrelation function, Grey Level Co-occurrence Matrix, Gabor filters, 3-level Haar wavelet transform, 3-level wavelet transform using 3-tap Daubechies filter and 3-level wavelet transform using 4-tap Daubechies filter) are evaluated and compared on a large corpus of historical documents in our previous work [24]. We concluded that the Gabor-based feature set is the best ones for font segmentation and for distinguishing textual regions from graphical ones. As a consequence, Gabor filters are applied on the enhanced DI (*cf.* Figure 2(c)) by using 4 different orientations ($\theta_g = \{0, \pi/4, \pi/2, 3\pi/4\}$) and 6 distinct spatial frequencies ($f_g = \{2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2}, 64\sqrt{2}\}$). An illustrative example of the magnitudes of 24 GFs, obtained by setting the 6 different spatial frequencies and 4 different orientations is presented in Figure 3.

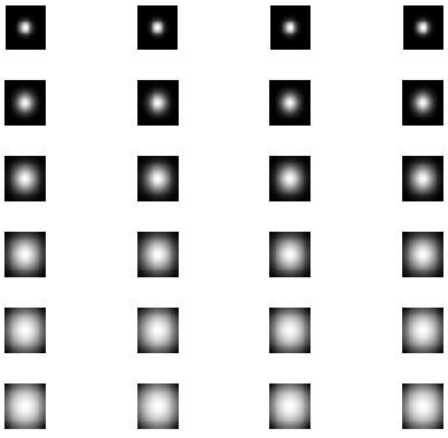


Figure 3: Illustration of the magnitudes of the specified Gabor filters by 6 different spatial frequencies f_g=$\{2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2}$ and $64\sqrt{2}\}$ and 4 different orientations θ_g=$\{0, \pi/4, \pi/2$ and $3\pi/4\}$.

Then, a quick and easy way to extract Gabor features on the whole transformed image by the selective Gabor filter, is to introduce a border replication step before the Gabor feature 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 technique.

Therefore, a Gabor-based 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. A 192-dimensional feature vector (48 Gabor indices $\times$ 4 sliding window sizes) is subsequently formed through the four different specified sizes of sliding windows.

2.3 Texture Feature Learning

Having the extracted Gabor features at the selected foreground superpixels, the superpixel classification is based on using a supervised machine learning (*i.e.* an SVM model) in order to discriminate the textual content from the graphical one. The supervised foreground superpixel classification task does not include spatial information. Using the computed Gabor feature vectors, the superpixel classification issue is modeled as a binary classification one (*i.e.* 1 for a textual content pixel while -1 for the graphical/noise content pixel). Therefore, two classes are defined, a class of text content pixels and another one containing all other contents such as noise (e.g. ink stains, mold or moisture, faded out ink, corrugated parchment or papyrus) and drawings (e.g. ornaments, drop caps, frames, embellishments, portraits). An SVM model using a linear kernel is trained to generate our classification model. To train an SVM model, the obtained 192-dimensional feature vectors of the selected foreground superpixels for every DI of the *"HBR2013 dataset"* are divided into two separate sets, namely, training set (60%) and testing set (40%). Subsequently, the training data of the *"HBR2013 dataset"* used to generate our classification model is obtained by combining the different training sets of a number of DIs selected randomly from the *"HBR2013 dataset"*. Given the training data, each marked for belonging to one of two classes (*i.e.* 1 or -1), our goal is to determine which class every foreground superpixel from the testing set in each HDI from the *"HBR2013 dataset"* will be in.

2.4 Superpixel Classification and Pixel Labeling

Since the texture feature learning phase has been performed, the selected foreground superpixels of the testing set are classified. Having classified foreground superpixels as textual or other contents such as noise and drawings based on the training data (*cf.* Section 2.3), a phase of labeling clusters of the foreground superpixels and pixels belonging to each superpixel in the enhanced HDI (*cf.* Figures 7(e), 7(f), 7(g) and 7(h)) is carried out with respect to the results of the superpixel classification phase. Since the superpixel classification and pixel labeling phases of the proposed algorithm have been performed, a pixel-labeled HDI is obtained (*cf.* Figures 7(q), 7(r), 7(s) and 7(t)).

3. EXPERIMENTS

We have experimentally evaluated the proposed ancient document enhancement and segmentation algorithm on 100 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.

3.1 Experimental Corpus and Ground Truth

In this section, a brief description of the experimental corpus and its associated ground truth is presented.

3.1.1 Experimental Corpus

Antonacopoulos *et al.* [3] considered a dataset as a good one if it is realistic (*i.e.* it must composed of real digitized DIs), comprehensive (*i.e.* it must well characterized and detailed for ensuring in-depth evaluation) and flexibly structured (*i.e.* to facilitate a selection of sub-sets with specific conditions).

Thus, in our experiments, we focus on real scanned HDIs. 100 images have been selected for historical document layout analysis and HBR competitions as part of the improving access to text (IMPACT)[2] project (an EU FP7 research project) and in the context of ICDAR conference and HIP workshop (2011 and 2013) [4, 5]. This dataset was called in this work the *"HBR2013 dataset"*. The *"HBR2013 dataset"* is composed of 100 binary, gray-scale or color HDIs which have been digitized at 150/300 dpi (*cf.* Figure 4). We have structured the *"HBR2013 dataset"* into two different categories differentiated by their content:

- 56 pages containing only text (*cf.* Figures 4(a), 4(b), 4(c) and 4(d)),
- 44 pages containing graphics and text (*cf.* Figures 4(e), 4(f), 4(g) and 4(h)).

The *"HBR2013 dataset"* which is used in different ICDAR competitions has the following characteristics: large variability of page content (*i.e.* document pages are different in writing and graphical style), complicated and complex page layout (e.g. several columns with irregular sizes, dense printing, irregular spacing, marginal notes), random alignment, use of specific and multiple fonts and illustration styles, large variability of editorial style and logical structure, presence of embellishments, irregular spacings (e.g. between characters, words, lines, paragraphs or margins), overlapping object boundaries, varying text column widths, interspersed graphics, frequent use of different kinds of graphics (e.g. ornaments, drop caps, frames, embellishments, portraits), noise and degradation caused by copying, scanning or aging (e.g. yellow pages, ink stains, mold or moisture, faded out ink, uneven lighting due to folded, corrugated parchment or papyrus), superimposition of information layers (e.g. stamps, handwritten notes at the margins, noise, back-to-front interference, ink that was bleeding through, historical spelling variants), page skew, scanning defects (e.g. curvature, light),

[2] `http://impact-project.eu`

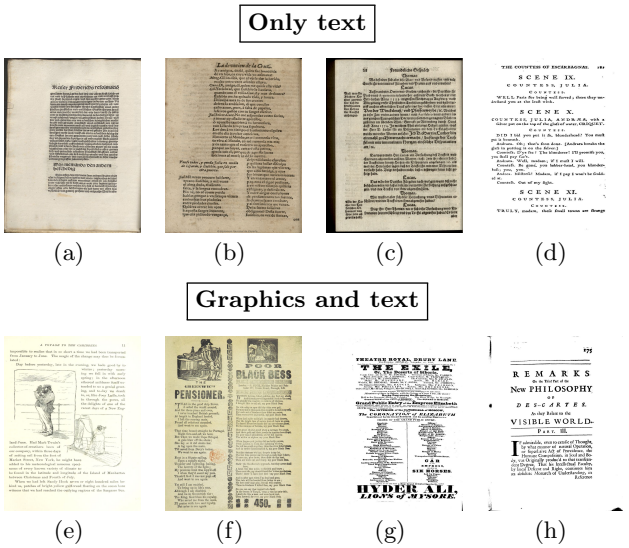


Figure 4: HDI examples of the *"HBR2013 dataset"*. Figures (a), (b), (c) and (d) illustrate few examples of HDIs containing only textual content. Figures (e), (f), (g) and (h) depict few examples of HDIs containing graphical and textual content.

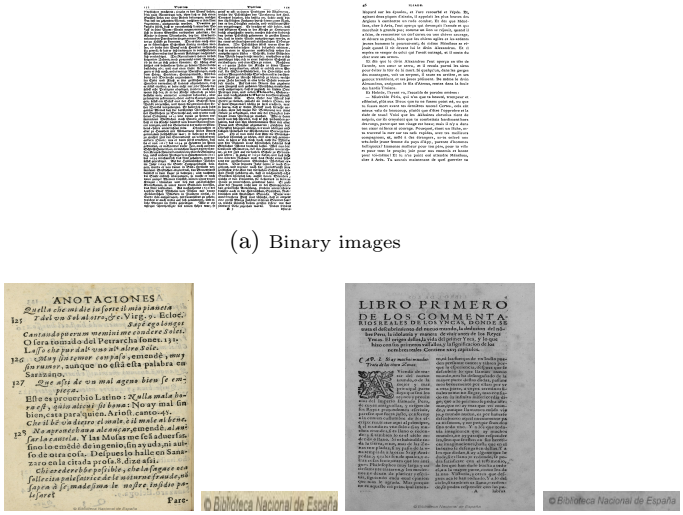


Figure 5: Illustration of the limitations of the *"HBR2013 dataset"*. Figure (a) shows few examples of binary images, while Figure (b) depicts few images of the *"HBR2013 dataset"* which have copyright notices at bottom of pages (zoomed regions on the copyright notices at bottom of pages are also illustrated).

presence of black borders, *etc.* It is composed of several binary images. Moreover, few images had been digitized at low resolution (*cf.* Figure 5(a)). Moreover, few images of the *"HBR2013 dataset"* have copyright notices at bottom of pages which may introduce an artificial information, thereafter inducing segmentation and characterization errors (*cf.* Figure 5(b)).

3.1.2 Ground Truth

Although the issues of the realistic dataset availability and broadband access to researchers for the performance evaluation of contemporary DIs have been discussed and solved by Antonacopoulos *et al.* [3], representative datasets of HDIs are still hard to collect from several libraries. Then, defining the associated ground truth of HDI corpus is still not a straightforward task due to the HDI characteristics (e.g. page skew, superimposition of information layers, such as stamps, handwritten notes, noise, back-to-front interference). These characteristics complicate the definition of the appropriate and objective ground truth, the characterization or segmentation of HDIs and make the processing of this kind of DIs a difficult task.

Our ground truth has been manually outlined using rectangular regions drawn around each selected zone. The regions have been ground truthed by zoning each content type (*i.e.* each rectangular region has been classified into text or graphics). Different labels for regions with different fonts have been also defined in order to evaluate the ability of texture features to separate various text fonts. Ground truth has been performed using the ground truthing editor, ground truthing environment for document images (GEDI)[3], a public domain DI annotation tool that labels spatial boundaries of regions [11]. By specifying rectangular regions on a DI 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 6). The ground truth has not been provided for all images of the *"HBR2013 dataset"* by the IMPACT research team (*i.e.* only six pages). Thus, the ground truth of the *"HBR2013 dataset"* has been also carried out by using the GEDI tool.

[3] `http://gedigroundtruth.sourceforge.net/`

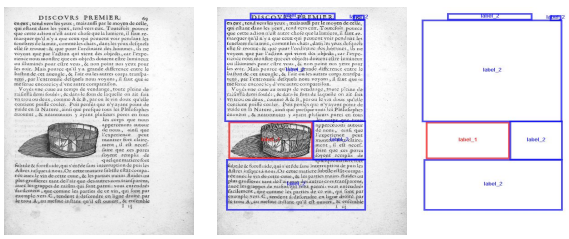


Figure 6: Example of the defined ground truth . Figure (a) illustrates an example of a HDI (as an input of the used domain DI annotation tool, GEDI). Figure (b) shows the specified and labeled rectangular regions using GEDI. The graphical regions have red labels ("label 1") while the textual ones have blue labels ("label 2").

3.2 Evaluation and Results

By visual inspection of the obtained results of the proposed enhancement and segmentation algorithm (*cf.* Figures 7(q), 7(r), 7(s) and 7(t)) for the HDIs from the *"HBR2013 dataset"* (*cf.* Section 3.1.1), we note that the presented method in

this article provides satisfying results particularly in distinguishing the textual regions (green) from the graphical ones (blue). In Figures 7(q), 7(r), 7(s) and 7(t), the selected foreground pixels representing textual content are labeled as green ones while those representing graphical and noise content are labeled as blue ones. We note in Figure 7(s) that the proposed method mis-classifies the horizontal net and considers as textual content (green). Moreover, in Figure 7(q), we show that few foreground pixels in the drop cap, which are located adjacent to textual content are mis-labeled (green). This confusion can be explained by the limitations of the Gabor-based method to separate spatially close distinct kinds of information (*i.e.* the vertical/horizontal spacing is too small). Indeed, the Gabor features are extracted for a specified range of frequency and direction values. Thus, the performance of a Gabor-based method depends directly on the layout document.

Then, in order to provide an additional analysis and get an insight into the classification accuracy, a confusion matrix, error matrix or contingency table (M_c) is computed through the analysis of the testing set of the selected foreground superpixels (*cf.* Table 1). From the M_c, several per-superpixel classification accuracy metrics, including precision (P), recall (R), F-score or F-measure (F) and classification accuracy rate (CA) are performed in this work.

- The ***precision metric*** (P) corresponds to the proportion of the predicted cases that are correctly matched to the benchmark classifications. It is considered as a means of assessing the classification in terms of false positives.
- The ***recall measure*** (R) indicates the proportion of real cases that are correctly predicted. It is considered a way to improve the classification.
- The ***F-measure*** (F) can be computed as a score resulting from the combination of the P and R accuracies by using a harmonic mean. It assesses both the homogeneity and the completeness criteria of a clustering result.
- The ***classification accuracy rate*** (CA) metric corresponds to the ratio of the true classified predicted pixels and the total number of pixels [15, 21].

By analyzing the confusion matrix illustrated in Table 1, whose elements represent the selected foreground superpixels, the individual class precision (P_i) and recall (R_i) are computed, where i denotes the investigated class. Precision is considered to be a means of assessing the classification while recall is considered as a way of improving the classification. We note that the graphical superpixels are classified with 47%(P) and 56%(R), while for the textual superpixels we find 98%(P) and 97%(R). Hence, the graphical class has lower precision and recall. Thus, we show that the proposed method tends to miss more graphical or noise superpixels than textual ones by labeling graphical or noise superpixels as belonging to the textual class. Thus, we note that the proposed method is more relevant for the segmentation and characterization of textual regions than graphical ones. This confirms that there are less training samples from graphical content. Moreover, the proposed method based on learning Gabor features is adequate for textual content segmentation, since the Gabor descriptors are known to be sensitive to the stroke width. In conclusion, the computed accuracy classification values are very promising (*cf.* Table 2). 73%, 77%, 75% and 96% of per-superpixel precision (P), recall (R), F-measure (F) and classification accuracy rate (CA) are noted, respectively.

Table 1: Evaluation of the proposed method of HDI enhancement and segmentation by calculating the confusion matrix.

		Ground truth: Graphic	Ground truth: Text	
Clustering outcomes	Class 1	458673	525333	$\leftrightarrow P_1 = 0.47$
	Class 2	366080	19326807	$\leftrightarrow P_2 = 0.98$
		$\updownarrow R_1 = 0.56$	$\updownarrow R_2 = 0.97$	

Table 2: Evaluation of the pixel labeling (*i.e.* text and graphical pixels) by computing several classification accuracy measures, precision (P), recall (R), F-measure (F) and classification accuracy rate (CA).

Accuracy metric	Value
Precision (P)	0.73
Recall (R)	0.77
F-measure (F)	0.75
Classification accuracy (CA)	0.96

4. DISCUSSION

The obtained results of the proposed method for enhancement and segmentation of HDIs are relevant. Nevertheless, the fundamental question is if the proposed method for enhancement and segmentation of HDIs has been assessed properly or not. We should point out that the main technological bottleneck is the definition of an accurate and objective ground truth. Antonacopoulos *et al.* [6] stated that a direct comparison between several algorithms is tough and critical task for a variety of DIA applications due to the need for a realistic data and the high requirement for an adequate ground truth as well as the use of a set of objective evaluation criteria. However, it is still hard to determine fairly the different HDI content types. An important issue can also be outlined which consists of the difficulty to take into account the noisy foreground class when defining the ground truth in the case of degraded HDIs. An and Baird [2] stipulated that the pixel-wise classifiers rely on the accuracy of ground truth annotations. Since the defined ground truth is not a pixel-based one (*i.e.* it is defined by spatial boundaries of regions with labels). This highlights the need for a pixel-based ground truth. This issue has been also reported by Kumar *et al.* [18] who outlined that the use of a zone-level ground truth might have an influence on the accuracy of pixel-level approach and particularly the recall measure.

In this work, the noise pixels have not been considered when defining our ground truth. To the best of our knowledge, there really is no defined pixel-based ground truth of HDIs which takes account the noise pixels. It is not a straightforward task to define appropriate and objective ground truth due to the characteristics of HDIs (e.g. page skew, superimposition of information layers, such as stamps, handwritten notes, noise, back-to-front interference). The first aspect of future work will be to use a new computer-aided ground truthing environment editor for creating and manipulating automatically meta-data corresponding to regions of interest on HDIs under consideration (*i.e.* to generate a pixel-based ground truth including the noise pixels). Then, our results will be improved if we include topographical or spatial relationships in our 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 classification results should be improved.

5. CONCLUSIONS

The proposed method aims at enhancing and segmenting the content of HDIs. Given that there are significant degradations and no hypothesis concerning the layout, the graphical properties or typographical parameters of the analyzed HDI, the use of a texture analysis technique for HDI has become an appropriate choice. The main idea of this article is to ensure a graphic/text segmentation in HDIs by extracting and analyzing texture features independently of the layout and content of the pages. A texture feature learning phase is integrated in the proposed method that the label of a foreground pixel is identified according to the textural characteristics of the different training sets of a number of HDI foreground pixels selected randomly from the *"HBR2013 dataset"*.

This work has shown the effectiveness of the proposed HDI enhancement and segmentation method on an experimental corpus which is composed of 100 HDIs. The experimental corpus is selected for historical document layout analysis and historical book recognition competitions in the context of ICDAR conference and HIP workshop. The proposed method is parameter-free and applicable to a large variety of HDIs. It does not assume *a priori* information regarding HDI layout and content. It is based on GFs at varying scale and an SVM classier trained on ground truth images that have been pre-segemented using SLIC superpixlels. Our future work will focusing on evaluating other classifiers such as RBF kernel and other superpixel techniques.

6. ACKNOWLEDGMENTS

This work was supported by the French national research agency (ANR), under Grant ANR-10-CORD-0020, which is gratefully acknowledged. The authors would like also to thank Christos Papadopoulos for providing access to the IMPACT dataset.

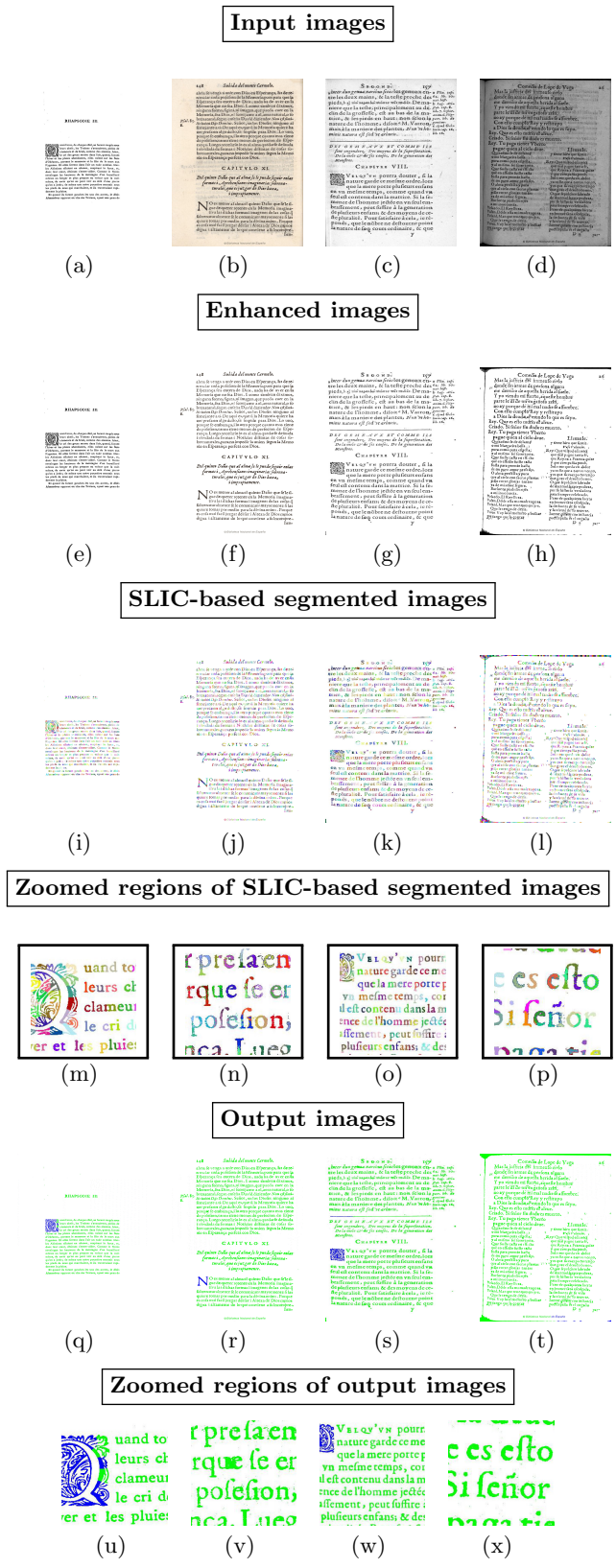


Figure 7: Examples of intermediate and resulting *"HBR2013 dataset"* images of the proposed method based on learning texture features for HDI enhancement and segmentation.

7. REFERENCES

[1] R. Achanta, A. Shaji, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence*, pages 2274–2282, 2012.

[2] C. An and H. S. Baird. The convergence of iterated classification. In *International Workshop on Document Analysis Systems*, pages 663–670. IEEE, 2008.

[3] A. Antonacopoulos, D. Bridson, C. Papadopoulos, and S. Pletschacher. A realistic dataset for performance evaluation of document layout analysis. In *International Conference on Document Analysis and Recognition*, pages 296–300. IEEE, 2009.

[4] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher. Historical document layout analysis competition. In *International Conference on Document Analysis and Recognition*, pages 1516–1520. IEEE, 2011.

[5] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher. ICDAR 2013 Competition on Historical Book Recognition (HBR 2013). In *International Conference on Document Analysis and Recognition*, pages 1459–1463. IEEE, 2013.

[6] A. Antonacopoulos, B. Gatos, and D. Bridson. Page segmentation competition. In *International Conference on Document Analysis and Recognition*, pages 1279–1283. IEEE, 2007.

[7] A. Asi, R. Cohen, K. Kedem, J. El-Sana, and I. Dinstein. A coarse-to-fine approach for layout analysis of ancient manuscripts. In *International Conference on Frontiers in Handwriting Recognition*, pages 140–145. IEEE, 2014.

[8] R. Cohen, A. Asi, K. Kedem, J. El-Sana, and I. Dinstein. Robust text and drawing segmentation algorithm for historical documents. In *International Workshop on Historical Document Imaging and Processing*, pages 110–117. ACM, 2013.

[9] A. Crasson and J. D. Fekete. Structuration des manuscrits : du corpus à la région. In *Colloque International Francophone sur l'Ecrit et le Document*, 2004.

[10] M. Diem and R. Sablatnig. Recognition of degraded handwritten characters using local features. In *International Conference on Document Analysis and Recognition*, pages 221–225. IEEE, 2009.

[11] D. Doermann, E. Zotkina, and H. Li. GEDI - a groundtruthing environment for document images. In *International Workshop on Document Analysis Systems*. ACM, 2010.

[12] A. Fischer, M. Baechler, A. Garz, M. Liwicki, and R. Ingold. A combined system for text line extraction and handwriting recognition in historical documents. In *International Workshop on Document Analysis Systems*, pages 71–75. IEEE, 2014.

[13] A. Garz, R. Sablatnig, and M. Diem. Layout analysis for historical manuscripts using SIFT features. In *International Conference on Document Analysis and Recognition*, pages 508–512. IEEE, 2011.

[14] C. Grana, G. Serra, M. Manfredi, D. Coppi, and R. Cucchiara. Layout analysis and content enrichment of digitized books. *Multimedia Tools and Applications*, pages 1–22, 2014.

[15] J. R. Jensen. *Introductory digital image processing*. Prentice-Hall, Englewood Cliffs, NJ, 1986.

[16] N. Journet, J. Ramel, R. Mullot, and V. Eglin. Document image characterization using a multiresolution analysis of the texture: application to old documents. *International Journal of Document Analysis and Recognition*, pages 9–18, 2008.

[17] K. Kise. *Page segmentation techniques in document analysis*. Handbook of Document Image Processing and Recognition, Springer-Verlag, 2014.

[18] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi. Text extraction and document image segmentation using matched wavelets and MRF model. *Image Processing*, pages 2117–2128, 2007.

[19] L. Likforman-Sulem. Apport du traitement des images à la numérisation des documents anciens. *Document Numérique*, pages 13–26, 2003.

[20] L. Likforman-Sulem, A. Zahour, and B. Taconet. Text line segmentation of historical documents: a survey. *International Journal of Document Analysis and Recognition*, pages 123–138, 2007.

[21] P. M. Mather. *Computer processing of remotely-sensed images: an introduction*. 2nd Edition John Wiley & Sons, 1999.

[22] M. Mehri, P. Gomez-Krämer, P. Héroux, A. Boucher, and R. Mullot. Texture feature evaluation for segmentation of historical document images. In *International Workshop on Historical Document Imaging and Processing*, pages 102–109. ACM, 2013.

[23] M. Mehri, V. C. Kieu, M. Mhiri, P. Héroux, P. Gomez-Krämer, M. A. Mahjoub, and R. Mullot. Robustness assessment of texture features for the segmentation of ancient documents. In *International Workshop on Document Analysis Systems*, pages 293–297. IEEE, 2014.

[24] M. Mehri, M. Mhiri, P. Héroux, P. Gomez-Krämer, M. A. Mahjoub, and R. Mullot. Performance evaluation and benchmarking of six texture-based feature sets for segmenting historical documents. In *International Conference on Pattern Recognition*, pages 2885–2890. IEEE, 2014.

[25] G. Nagy. Twenty years of document image analysis in PAMI. *Pattern Analysis and Machine Intelligence*, pages 38–62, 2000.

[26] O. Okun and M. Pietikäinen. A survey of texture-based methods for document layout analysis. In *Workshop on Texture Analysis in Machine Vision*, pages 137–148. Springer-Verlag, 1999.

[27] S. Uttama, P. Loonis, M. Delalandre, and J. M. Ogier. Segmentation and retrieval of ancient graphic documents. In *International Workshop on Graphics Recognition*, pages 88–98. Springer-Verlag, 2006.

[28] H. Wei, M. Baechler, F. Slimane, and R. Ingold. Evaluation of SVM, MLP and GMM classifiers for layout analysis of historical documents. In *International Conference on Document Analysis and Recognition*, pages 1252–1256. IEEE, 2013.