Extraction of Homogeneous Regions in Historical Document Images

Maroua Mehri[1,2], Pierre Héroux[2], Nabil Sliti[3], Petra Gomez-Krämer[1],
Najoua Essoukri Ben Amara[3] and Rémy Mullot[2]
[1]*L3i, University of La Rochelle, Avenue Michel Crépeau, 17042, La Rochelle, France*
[2]*LITIS, University of Rouen, Avenue de l'Université, 76800, Saint-Etienne-du-Rouvray, France*
[3]*SAGE, University of Sousse, École Nationale d'Ingénieurs de Sousse, 4023, Sousse, Tunisia*

Keywords: Historical Document Images, Segmentation, SLIC Superpixels, Gabor Filters, Multi-Scale Analysis, ARLSA.

Abstract: To reach the objective of ensuring the indexing and retrieval of digitized resources and offering a structured access to large sets of cultural heritage documents, a raising interest to historical document image segmentation has been generated. In fact, there is a real need for automatic algorithms ensuring the identification of homogenous regions or similar groups of pixels sharing some visual characteristics from historical documents (*i.e.* distinguishing graphic types, segmenting graphical regions from textual ones, and discriminating text in a variety of situations of different fonts and scales). Indeed, determining graphic regions can help to segment and analyze the graphical part in historical heritage, while finding text zones can be used as a pre-processing stage for character recognition, text line extraction, handwriting recognition, *etc.* Thus, we propose in this article an automatic segmentation method for historical document images based on extraction of homogeneous or similar content regions. The proposed algorithm is based on using simple linear iterative clustering (SLIC) superpixels, Gabor filters, multi-scale analysis, majority voting technique, connected component analysis, color layer separation, and an adaptive run-length smoothing algorithm (ARLSA). It has been evaluated on 1000 pages of historical documents and achieved interesting results.

1 INTRODUCTION

The idea of conducting strategies of digitization programs with cultural heritage documents has emerged since the early 1960s. The primary goals of these digitization programs, which were related to the tremendous growth and spread of the Internet technologies, were not clearly identified (*e.g.* providing digital copies of historical document images (HDIs), sharing databases of document images between many libraries, designing a computer-assistance tool for textual data handling, *etc.*). Nevertheless, the rapid growth of digital libraries has become a serious hindrance to promote wide efficiency and effectiveness in the management of this cultural heritage resources (*i.e.* quick and relevant access to information contained therein) due to the huge amount of digital high quality reproductions of fragile books and digital copies of rare collections. To meet the need to reinforce the enrichment and exploitation of heritage documents in addition to make it electronically available for access via the Internet, many research projects has been set up with the support of public funding provided through the European and American governments. The main goals of these projects are to provide a computer-based access and analysis of cultural heritage documents, searchable and browseable HDI databases, and an automatic indexing, linking and retrieval semantic-based systems of HDIs (Coustaty et al., 2011).

In this work, we are interested in historical document image layout analysis (HDILA). HDILA starts by segmenting a document in order to find and classify homogeneous regions or zones, such as graphic and textual regions (Okun and Pietikäinen, 1999). Finding graphic regions can be used to analyze the graphical part in historical heritage, while determining text zones can be used as a pre-processing stage for handwriting recognition, *etc.* Our goal consists of identifying homogenous regions or similar groups of pixels sharing some visual characteristics by labeling and grouping pixels from HDIs. We aim to characterize each digitized page of historical book with a set of homogeneous or similar content regions and

Mehri M., Héroux P., Sliti N., Gomez-Krämer P., Ben Amara N. and Mullot R..
Extraction of Homogeneous Regions in Historical Document Images.
DOI: 10.5220/0005265500470054
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 47-54
ISBN: 978-989-758-091-8

their topological relationships to characterize the document layout and content by defining one or more signatures for each digitized page. Therefore, an automatic segmentation method for HDIs is proposed in this article. The proposed algorithm is based on using simple linear iterative clustering (SLIC) superpixels (Liu et al., 2011), Gabor filters (Gabor, 1946), k-means clustering (MacQueen, 1967), multi-scale analysis (Li et al., 2000), majority voting technique (Lam and Suen, 1997), connected component (CC) analysis (Rosenfeld and Pfaltz, 1966), color layer separation, and adaptive run-length smoothing algorithm (ARLSA), for extraction of homogeneous or similar content regions in HDIs.

The remainder of this article is organized as follows: in Section 2, the proposed segmentation algorithm for HDIs based on extraction of homogeneous or similar content regions is detailed. Section 3 describes the experimental protocol by presenting the experimental corpus, and the defined ground-truth. To evaluate the performance of our proposed algorithm and validate our choice of the used techniques on each step of our algorithm, a set of experiments on a large variety of HDIs is detailed in Section 4. Then, an assessment of the different steps of our algorithm is presented and an analysis of the obtained results is subsequently discussed. Qualitative results are also given to demonstrate the segmentation quality. Our conclusions and future work are presented in Section 5.

2 PROPOSED METHOD

To ensure a relevant segmentation of graphical regions from textual ones, an efficient discrimination of text in a variety of situations of different fonts and scales, and a robust distinction between different graphic types in HDIs, an automatic segmentation method for HDIs based on extraction of homogeneous or similar content regions is proposed in this article. The proposed algorithm segments the content of digitized HDIs. In particular, it discriminates between the different classes of the foreground layers of a digitized document based on textural and topological descriptors. First, a HDI is fed as input and is read as a gray-scale image. The extraction of texture information 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 textural information. Then, to obtain enhanced backgrounds of noisy HDIs and reduce the step complexity of a pixel-based segmentation method, a foreground-background segmentation task based on SLIC superpixel segmentation and k-means clustering algorithms is performed. Afterwards, an automatic extraction of texture descriptors from foreground superpixels is processed by involving a multi-scale approach. The extracted textural features are then used in an unsupervised clustering approach to label clusters or groups of pixels with respect to the results of the superpixel clustering phase. Then, for refinement of the pixel labeling results, a first step of post-processing *"Post-processing 1"* is introduced by taking into consideration the topological relationship between pixels and integrating a spatial multi-scale analysis of majority votes. Finally, a second post-processing task *"Post-processing 2"* is added based on using the multi-scale analysis, majority voting technique, CC analysis, color layer separation, and ARLSA to identify the homogeneous or similar content regions that are characterized by similar properties.

The proposed algorithm does not require *a priori* knowledge of the document structure/layout, the typographical parameters or the graphical properties of the document image. In addition, it is fast since a SLIC superpixel technique has been used in our proposed algorithm, 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 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 (Achanta et al., 2012). Figure 1 illustrates the detailed schematic block representation of the proposed algorithm.

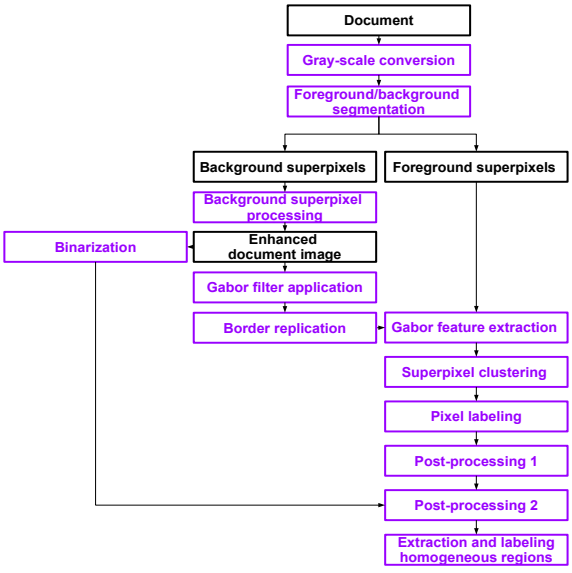

Figure 1: Flowchart of the proposed algorithm for extraction of homogeneous regions in HDIs.

2.1 Foreground-background Segmentation

By using the SLIC superpixel approach on our proposed algorithm, pixels sharing similar characteristics or properties (*e.g.* texture cues, contour, color, *etc.*) are grouped into a significant polygon-shaped region (Achanta et al., 2012). Thus, by setting the number of superpixels k_s equal to 0.01% of image pixels, an over-segmented image representing a compact content map is generated. 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. To segment an image into two layers (*i.e.*, the foreground and background), 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. One represents the information of the background (*cf.* Figure 3(b)) and the other represents the foreground (*e.g.* noise, text fields, drawings, *etc.*) (*cf.* Figure 3(c)).

2.2 Document Enhancement

Since the foreground-background 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.* 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 3(d)). Figure 3(d) illustrates an example of enhanced image by the superpixel technique with a clean background.

2.3 Gabor Feature Extraction

In our previous work, some of the well-known texture-based approaches (auto-correlation function, gray-level co-occurrence matrix, and Gabor filters) were compared for ancient document image segmentation (Mehri et al., 2013). We concluded that the Gabor features perform better than the auto-correlation and co-occurrence ones for font segmentation and for distinguishing textual regions from graphical ones. Thus, once the document is enhanced, Gabor filters are applied on it by setting the same default parameters proposed in (Mehri et al., 2013). 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 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. A 192-dimensional feature vector (48 Gabor indices $\times$ 4 sliding window sizes) is subsequently formed through four different sizes of sliding windows.

2.4 Foreground Superpixel Clustering and Foreground Pixel Labeling

A foreground superpixel clustering task is performed by partitioning Gabor-based feature sets into compact and well-separated clusters in the feature space to ensure the segmentation of graphical regions from textual ones, the discrimination text in a variety of situations of different fonts and scales, and the distinction between different graphic types in HDIs. The foreground superpixel clustering task does not include spatial information and is performed by using the k-means algorithm. Then, 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 have been performed, a pixel-labeled document image is obtained (*cf.* Figure 3(f)).

2.5 Post-processing 1

To refine the pixel labeling results, many researchers have introduced the spatial relationships between pixels which have not been considered when the texture features have been analyzed. Then, it has the advantage to deal with the non-smoothed boundaries due to the extraction of texture features from small predefined windows (Chang and Kuo, 1992). In our algorithm, a first step of post-processing *"Post-processing 1"* is introduced by taking into consideration the topological relationship between the selected foreground superpixels and integrating a spatial multi-scale analysis of majority votes. First, the Euclidean distance

between each foreground superpixel and the centroid of cluster belonging to it is computed. Then, the foreground superpixels are sorted in descending order according to the computed Euclidean distance values in such a way that the first processed foreground superpixel is the one that has a higher Euclidean distance value. The higher the values of the computed Euclidean distances, the more there is a high probability that the foreground superpixel is improperly labeled since it is far from the centroid of cluster belonging to it. Thus, the first processed foreground superpixels are those that have high values of Euclidean distances by using a multi-scale majority voting technique. By performing a multi-scale approach in the majority voting technique, small isolated groups of superpixels will be removed. Indeed, a local decision on the label of each selected foreground superpixel is taken using the maximum number or majority of superpixel labels and pixel labels belonging to it, which is performed at the same four pre-defined sizes of sliding windows in the Gabor feature extraction step. Then, if the processed foreground superpixel has a new label, the pixels belonging to it will have the same new label. Afterwards, the next processed foreground superpixel is one that has a smaller Euclidean distance value than the former foreground superpixel. The labels of foreground superpixels and the pixels belonging to them are updated on each run of multi-scale majority voting technique on each foreground superpixel to ensure a relevant refinement of the pixel labeling results. Since the first step of post-processing of the proposed algorithm *"Post-processing 1"* has been performed, a post-processed 1 pixel-labeled document image is obtained (*cf.* Figure 3(g)).

2.6 Post-processing 2

As already seen on the proposed algorithm (*cf.* Figure 1), our goal is to find homogeneous regions defined by common characteristics or similar texture features as easily, quickly, and automatically as possible. So since the first step of post-processing *"Post-processing 1"* has been performed, our output data consists of a post-processed 1 pixel-labeled document image. Nevertheless, we need to identify group of pixels sharing common characteristics or similar textural properties at this stage in order to extract homogenous region (*i.e.* to partition text into columns, paragraphs, lines or words, and identify the graphical regions). Therefore, we aim in the second step of post-processing *"Post-processing 2"* to fill automatically the space within each pixel in order to determine the largest CCs illustrating similar content regions by replacing a sequence of background pixels with foreground ones and afterwards grouping pixels which share common characteristics or similar textural properties from the post-processed 1 pixel-labeled document image (*cf.* Section 2.5, Figure 3(g)).

First, a binarization step is performed using a standard parameter-free binarization method, the Otsu's method, on the enhanced document image (*cf.* Section 2.2, Figure 3(d)) to obtain a binarized enhanced document image (*cf.* Figure 3(e)) and subsequently to retrieve the CCs (Otsu, 1979). Then, the majority voting technique is applied on each extracted CC from the binarized enhanced document image by computing the maximum number or majority of pixel labels belonging to the localized CC on the post-processed 1 pixel-labeled document image (*cf.* Section 2.5, Figure 3(g)). Therefore, using the majority voting technique, the extracted CCs from the binarized enhanced document image are labeled according to the post-processed 1 pixel-labeled document image. The resulting image of labeling the extracted CCs is illustrated in Figure 3(h).

Since the extracted CCs from the binarized enhanced document image are labeled, a color layer separation task is performed to split the CCs according to their labels. Therefore, a document image containing only single color CCs is generated for each color layer. For instance, in the example illustrated in Figure 3, there are two colors representing separately the graphical (blue) and textual (green) CCs in Figures 3(i) and 3(m), respectively. The color layer separation task ensures the segmentation of the extracted CCs according to their label (*i.e.* content type). When we separate the extracted CCs according to their label, the issues caused by the complex, dense, and overlapping document layout of HDIs will be overcome. The identification of homogeneous regions is based on finding the largest CCs. By replacing a sequence of background pixels with foreground ones and afterwards grouping pixels which share common characteristics or similar textural properties from a pixel-labeled document image, the extraction of homogeneous regions will be more accurate and relevant. Indeed, the idea is to fill automatically the space within each component to partition text into columns, paragraphs, lines or words on the one hand, and identify the graphical regions on the other hand.

So an adaptive RLSA is proposed in this work, which is a modified version of the state-of-the-art RLSA (Wahl et al., 1982). The RLSA studies the spaces between black pixels in order to link neighboring black areas by applying the run-length smearing both horizontally and vertically. It operates by replacing a horizontal (vertical, respectively) sequence of background pixels with foreground ones if the num-

ber of background pixels in the horizontal (vertical, respectively) sequence is smaller or equal to a predefined horizontal (vertical, respectively) threshold. The proposed ARLSA determines automatically the horizontal and vertical thresholds, which correspond to to the run-length smoothing values, respectively. To obtain the proper values of the horizontal and vertical thresholds, the two histograms of the widths and heights of the extracted CCs are examined, respectively. These two histograms gives the distributions of the widths and heights of the extracted CCs in the analyzed HDI. The estimation of the horizontal (vertical, respectively) threshold is based on the determination of the global maximum of the histogram of the widths (heights, respectively) of the extracted CCs. The global maximum of the histogram of the widths (heights, respectively) of the extracted CCs gives mainly information about the mean character length (height, respectively).

Once the horizontal and vertical run-length smoothing values are estimated automatically according to the analyzed HDI content (*i.e.* and particularly the distributions of the widths and heights of the extracted CCs of the binarized enhanced document image), the proposed ARLSA is applied on each resulting image of the color layer separation task after performing a binarizing step by using the Otsu's algorithm. It operates by taking the logical AND of the horizontally (*cf.* Figure 3(j) (3(n), respectively)) and vertically (*cf.* Figure 3(k) (3(o), respectively)) merged images of each resulting image of the color layer separation task to generate Figure 3(l) (3(p), respectively). After applying the ARLSA on each resulting image of the color layer separation task, the logical NOT is performed on each resulting image to merge the different resulting images of the ARLSA task (*cf.* Figures 3(l) and 3(p)) with the logical OR. Since the merge process of the different resulting images has been performed with the logical OR, a post-processed binarized document image is generated (*cf.* Figure 3(q)) in which the neighboring black areas are linked by applying the run-length smearing both horizontally and vertically. Then, the post-processed 2 pixel-labeled document image (*cf.* Figure 3(r)) is obtained with labeling the extracted CCs from Figure 3(q), according to the deduced labels from Figure 3(h) by using the majority voting technique.

Finally, the homogeneous or similar content regions are extracted and labeled from the resulting image of the *"Post-processing 2"* task by identifying group of pixels sharing common characteristics or similar textural properties (*cf.* Figure 3(s)). To define an extracted region, a bounding box covering all the pixels belonging to the extracted CC is used (*i.e.* a contour tracking of the shape of the extracted CC is carried out to identify the bounding box from each component). Then, the colors of the external contours of the defined bounding box is drawn according to the label deduced from the resulting image of using the majority voting technique (*cf.* Figure 3(h)).

3 CORPUS AND PREPARATION OF GROUND-TRUTH

Our experimental corpus contains 1000 ground-truthed one-page document images which have been collected from Gallica, encompassing six centuries (1200-1900) of French history. The HDIs of our corpus have been selected from several printed monographs and manuscripts across a variety of disciplines (*e.g.* novels, law texts, educational books (history, geography, nature), xylographic booklets, *etc.*) to provide a broader range of document contents. They are gray-scale/color documents which have been digitized at 300/400 dpi and saved in the "TIFF" format, which provides a high resolution of digitized images.

Our dataset has been structured into four categories of real scanned HDIs differentiated by their content (*cf.* Figure 2), reflecting the challenges of our work to determine which texture features can be more adequate for segmenting 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 HDIs with both simple and complex layouts for each category of documents which have been ground-truthed to ensure a better understanding of the behavior of the evaluated texture feature sets. It is composed of:

- 250 pages containing only two fonts (*cf.* Figure 2(a))
- 250 pages containing only three fonts (*cf.* Figure 2(b))
- 250 pages containing graphics and one text font (*cf.* Figure 2(c))
- 250 pages containing graphics and text with two different fonts (*cf.* Figure 2(d))

(a) Only two fonts (b) Only three fonts (c) Graphics and one text font (d) Graphics and two text fonts

Figure 2: Illustration of our experimental corpus.

The ground-truth for document images has been manually outlined using rectangular regions drawn around each selected zone. The zones have been ground-truthed by zoning manually 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 for evaluating the performance of texture feature to separate various text fonts.

4 EVALUATION AND RESULTS

In this work, due to a possible bias produced by estimating the number of clusters, the maximum number of homogeneous and similar content regions is set equal to the ones defined in the ground-truth. The first aspect of future work is to introduce a clustering methodology 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).

The proposed algorithm provides very satisfying results particularly in distinguishing textual regions from graphical ones (*cf.* Figure 3(s)). This highlights a much greater discriminant power for separating text and graphic regions than for distinguishing two or more different text fonts (normal, uppercase, and italic fonts). Nevertheless, this way of assessing the effectiveness of a segmentation method is inherently a subjective evaluation and we need to evaluate robustness using an appropriate quantitative metric. In the tables there are two *"Overall"* values. The *"Overall***"* value is obtained by averaging all the respective column values except the value of *"Two fonts and graphics****"*. The *"Overall****"* value is obtained by averaging all the respective column values except the value of *"Two fonts and graphics***"*. *"Two fonts and graphics***"* represents the case when every font in the text has a different label in the ground truth, and clustering is performed by setting the number of types of content regions to 3 (graphics and two different text fonts). *"Two fonts and graphics****"* represents the case when all fonts in the text have the same label in the ground truth, and clustering is performed by setting the number of types of content regions equal to 2 (graphics and text). This distribution of this dataset points out whether or not the proposed algorithm is firstly adequate for segmenting graphical regions from textual ones, and secondly if it can discriminate between texts with a variety of fonts and scales. The results of accuracy metrics are presented in Table 1.

First, in this study the F-measure (F) is computed to evaluate the different steps of post-processing of the proposed algorithm for extraction of homogeneous regions in HDIs. The overall pixel labeling results are reasonably promising, *i.e.* we obtain 67% and 70% of overall F-measure rates, without taking into consideration the topological relationships of pixels and their labels for *"Overall***"* and *"Overall****"*, respectively. We conclude that the best performance is obtained for documents containing graphics and single text font (81%). The lowest value of F is obtained for documents containing only three fonts (53%). Therefore, the pixel labeling results show a much greater discriminating power for separating text (single font) and graphic regions than for distinguishing documents containing graphics and two or more text fonts. Furthermore, we note that we do not need a post-processing phase, since there is a no significant performance difference between the cases of without and with adding the multi-scale analysis of majority voting (*i.e.* overall F gains of 0.1% and 0.004% for *"Overall***"* and *"Overall****"*, respectively). This study shows the robustness of the proposed pixel labeling technique based on Gabor feature analysis with the SLIC superpixels and multi-scale approach for segmentation in the case of the uselessness of introducing the topological relationships. However, by adding a post-processing step to identify group of pixels sharing common characteristics or similar textural properties and extract homogenous region, overall F gains of 0.6% and 1% are obtained for *"Overall***"* and *"Overall****"*, respectively.

Then, three accuracy metrics are computed: the precision (P_{AR}), recall (R_{AR}), and Jaccard index (J_{AR}) for evaluating the extracted homogeneous regions (Brunessaux et al., 2014). The results obtained for homogenous region extraction assessment are performed (P_{AR}, R_{AR}, and J_{AR}) by calculating the mean value of each accuracy metric relative to the number of defined rectangular regions of the ground-truth, which are illustrated in Table 1. The computed accuracy metrics values are quite encouraging since we obtain 88%(P_{AR}), 90%(R_{AR}), and 86%(J_{AR}). 93%(P_{AR}), 93%(R_{AR}), and 91%(J_{AR}) are noted for documents containing text (single font) and graphic regions. In conclusion, by computing numerous accuracy metrics, we prove that the proposed algorithm has a much greater discriminating power for separating text (single font) and graphic regions than for distinguishing documents containing graphics and two or more text fonts. The results also confirm that it is more difficult to separate two or three text fonts in HDIs.

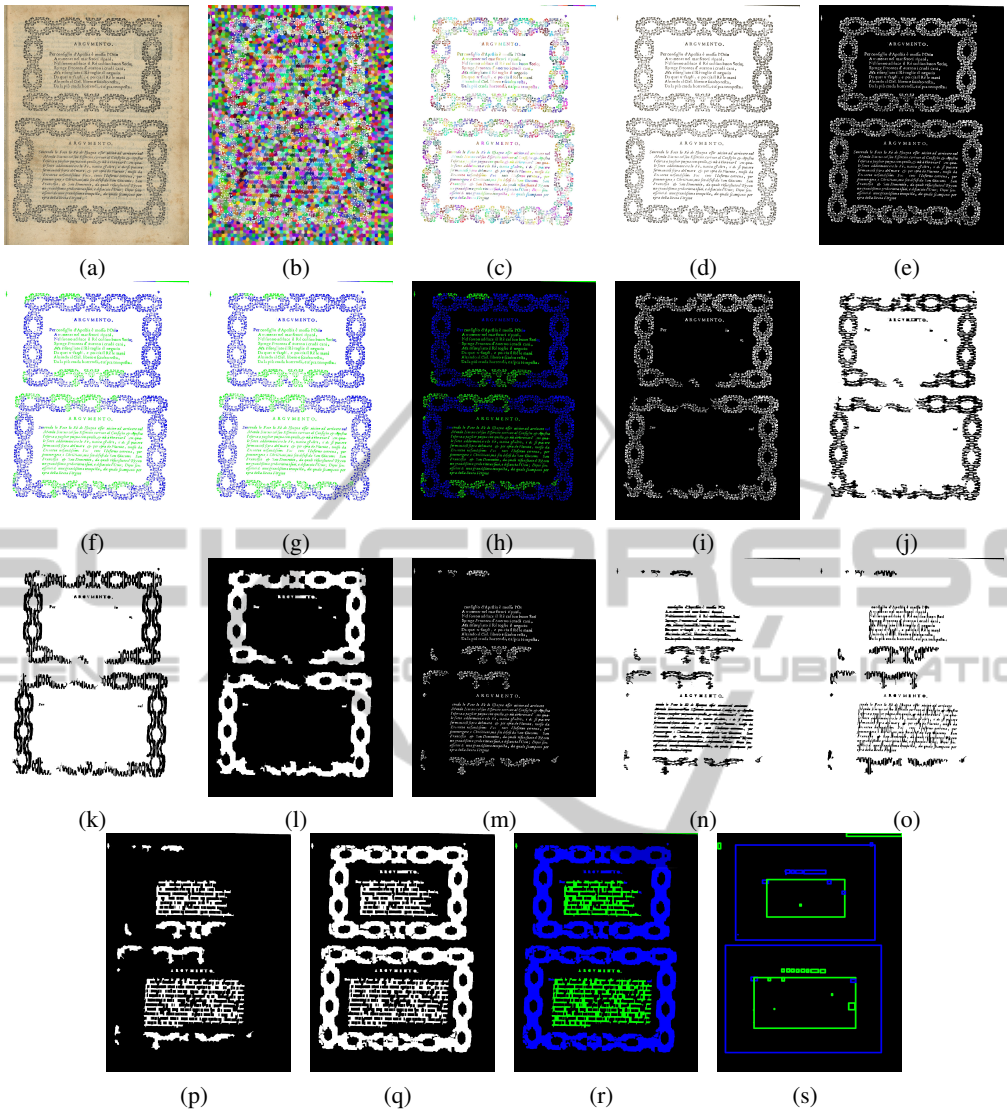

Figure 3: Illustration of the intermediate results of the different tasks of the post-processing step of the proposed algorithm: Figure (a) shows an example of an HDI (as an input of the proposed algorithm). Figures (b) and (c) show the background and foreground SLIC superpixels, respectively. Colors assigned to the background (foreground, respectively) superpixels which are illustrated in Figure (b) ((c), respectively) are randomly generated. Figure (d) depicts the result of the enhancement step. Figure (e) shows the resulting image of applying a binarization step on the enhanced document image. Figure (f) illustrates the pixel-labeled image (as an output of the analysis of the extracted Gabor features (graphic regions (blue), textual regions (green)). Figure (g) depicts the results of the first step of post-processing *"Post-processing 1"*. Figure (h) shows the resulting image of the labeling of the extracted CCs from the binarized enhanced document image according to the obtained pixel labeling results in the first step of post-processing by using the majority voting technique. Figures (i) and (m) are the two resulting binarized images of the color layer separation task, illustrating separately the graphical (blue) and textual (green) CCs, respectively. Figures (j) and (k) show the resulting images of the application of the run-length smearing both horizontally and vertically on the resulting binarized image representing the graphical regions (*cf.* Figure (i)), respectively. Figures (n) and (o) show the resulting images of the application of the run-length smearing both horizontally and vertically on the resulting binarized image representing the textual regions (*cf.* Figure (m)), respectively. Figure (l) ((p), respectively) is the resulting image of merging the two resulting images of applying the run-length smearing both horizontally and vertically on each resulting binarized image of the color layer separation task (*cf.* Figures (j) and (k)) (*cf.* Figures (n) and (o), respectively) by using the logical AND. Figure (q) is the resulting image of merging the two resulting images of applying ARLSA on each resulting binarized image of the color layer separation task by using the logical OR (*cf.* Figures l) and (p), respectively). Figure (r) shows the resulting image of the second step of post-processing *"Post-processing 2"* by labeling the extracted CCs from Figure (q) with taking into account to the labels of the extracted CCs from Figure (h). Figure (s) illustrates the output of the proposed algorithm for extraction of homogeneous regions in HDIs.

Table 1: Evaluation of the different steps of the proposed algorithm for extraction of homogeneous regions in HDIs by computing the F-measure (F), the precision (P_{AR}), the recall (R_{AR}), and the Jaccard index (J). $\mu(.)$ represents the mean value. The higher the mean values, the better the results. †, ‡, and ≀ represent the evaluation of the analysis of the extracted texture features in the case of without the first step of post-processing "the pixel labeling step", with the first step of post-processing and with the second step of post-processing, respectively. $\mu^{\ddagger-\dagger}(.)$ and $\mu^{\wr-\ddagger}(.)$ represent the mean difference values between ‡ and †, and between ≀ and ‡, respectively.

	$\mu^{\dagger}(F)$	$\mu^{\ddagger}(F)$	$\mu^{\wr}(F)$	$\mu^{\ddagger-\dagger}(F)$	$\mu^{\wr-\ddagger}(F)$	$\mu(P_{AR})$	$\sigma(R_{AR})$	$\mu(J_{AR})$
One font and graphics	0.8159	0.8172	0.7761	0.0013	–0.0557	0.9264	0.9352	0.9158
Two fonts and graphics*	0.6011	0.6019	0.6313	0.0008	0.0294	0.9158	0.9282	0.9071
Two fonts and graphics**	0.7140	0.7125	0.7553	–0.0015	0.0428	0.8223	0.8851	0.8429
Only two fonts	0.7315	0.7322	0.7403	0.0007	0.0081	0.8227	0.8782	0.8373
Only three fonts	0.5342	0.5356	0.5816	0.0014	0.0460	0.8628	0.8564	0.7970
Overall*	0.6706	0.6717	0.6786	0.0010	0.0069	0.8818	0.9012	0.8657
Overall**	0.6989	0.6993	0.7096	0.0004	0.0103	0.8819	0.8995	0.8643

5 CONCLUSION AND PERSPECTIVES

This article presents a novel automatic segmentation method for HDIs based on extraction of homogeneous or similar content regions. The proposed algorithm is based on using simple linear iterative clustering (SLIC) superpixels, Gabor filters, multi-scale analysis, majority voting technique, CC analysis, color layer separation, and ARLSA. The robustness of the proposed algorithm is used in a parameter-free method and adapted to all kinds of HDIs which is designed to identify the homogeneous regions without formulating a hypothesis or assumption concerning the document model/layout or content. It was evaluated on 1000 pages of HDIs with promising results.

Homogeneous region extraction in HDIs is a first step towards automatic historical book understanding, our future work will build on the results of the extraction of homogeneous or similar content regions to characterize HDI content with intermediate level meta-data, between document content and layout. By characterizing each digitized page of historical book with a set of homogeneous or similar content regions and their topological relationships, a signature can be designed for each book page. The obtained signatures will help deducing the similarities of book page structure or layout and/or content.

REFERENCES

Achanta, R., Shaji, A., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *PAMI*, pages 2274–2282.

Brunessaux, S., Giroux, P., Grilheres, B., Manta, M., Bodin, M., Choukri, K., Galibert, O., and Kahn, J. (2014). The Maurdor project: Improving automatic processing of digital documents. In *DAS*, pages 349–354. IEEE.

Chang, T. and Kuo, C. C. J. (1992). Texture segmentation with tree-structured wavelet transform. In *TFTSA*, pages 543–546.

Coustaty, M., Raveaux, R., and Ogier, J. M. (2011). Historical document analysis: A review of French projects and open issues. In *EUSIPCO*, pages 1445–1449.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, pages 429–441.

Lam, L. and Suen, C. Y. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *SMC*, pages 553–568.

Li, J., Wang, J. Z., and Wiederhold, G. (2000). Classification of textured and non-textured images using region segmentation. *IP*, pages 754–757.

Liu, M. Y., Tuzel, O., Ramalingam, S., and Chellappa, R. (2011). Entropy rate superpixel segmentation. In *CVPR*, pages 2097–2104.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. University of California Press.

Mehri, M., Gomez-Krämer, P., Héroux, P., Boucher, A., and Mullot, R. (2013). Texture feature evaluation for segmentation of historical document images. In *HIP*, pages 102–109.

Okun, O. and Pietikäinen, M. (1999). A survey of texture-based methods for document layout analysis. In *WTAMV*, pages 137–148.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *SMC*, pages 62–66.

Rosenfeld, A. and Pfaltz, J. L. (1966). Sequential operations in digital picture processing. *Journal of the ACM*, pages 471–494.

Wahl, F. M., Wong, K. Y., and Casey, R. G. (1982). Block segmentation and text extraction in mixed text/image documents. *CGIP*, pages 375–390.