A Bottom-up Method Using Texture Features and a Graph-based Representation for Lettrine Recognition and Classification

Maroua Mehri*†, Petra Gomez-Krämer*, Pierre Héroux†, Mickaël Coustaty*, Julien Lerouge†, and Rémy Mullot*

*L3i, University of La Rochelle, Avenue Michel Crépeau, 17042, La Rochelle, France
Emails: {maroua.mehri, mickael.coustaty, petra.gomez, remy.mullot}@univ-lr.fr
†LITIS, University of Rouen, Avenue de l'Université, 76800, Saint-Etienne-du-Rouvray, France
Emails: {pierre.heroux, julien.lerouge}@univ-rouen.fr

Abstract—This article tackles some important issues relating to the analysis of a particular case of complex ancient graphic images, called "lettrines", "drop caps" or "ornamental letters". Our contribution focuses on proposing generic solutions for lettrine recognition and classification. Firstly, we propose a bottom-up segmentation method, based on texture, ensuring the separation of the letter from the elements of the background in an ornamental letter. Secondly, a structural representation is proposed for characterizing a lettrine. This structural representation is based on filtering automatically relevant information by extracting representative homogeneous regions from a lettrine to generate a graph-based signature. The proposed signature provides a rich and holistic description of the lettrine style by integrating varying low-level features (e.g. texture). Then, to categorize and classify lettrines with similar style, structure (*i.e.* ornamental background) and content (*i.e.* letter), a graph-matching paradigm has been carried out to compare and classify the resulting graph-based signatures. Finally, to demonstrate the robustness of the proposed solutions and provide additional insights into their accuracies, an experimental evaluation has been conducted using a relevant set of lettrine images. In addition, we compare the results achieved with those obtained using the state-of-the-art methods to illustrate the effectiveness of the proposed solutions.

Keywords—Lettrine, Recognition, Classification, Indexing, Texture, Graph.

I. INTRODUCTION

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 have been set up with the support of public funding provided by 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 historical document image databases, and an automatic indexing, linking and retrieval semantic-based systems of historical document images. Some works have been proposed to deal with the whole historical document images or digitized ancient books [1], while others have been focused on investigating and analyzing parts of historical document images (e.g. text, styles/fonts, handwritten annotations) [2] or graphic images (e.g. illustrations and drop caps [3]). Historical document images often contain graphical features represented by lettrines (*cf.* Figure 1). Lettrines were widely present in digitized ancient books of the 15th and 16th centuries. They

occur at the beginning of a chapter or paragraph. Other terms are often used to define a lettrine such as "drop cap", "drop capital" or "ornamental letter". A lettrine is usually represented by two main layers, letter or initial and drawing painted in the background. As a matter of fact, the lettrines can be composed of three principle components: letter, pattern and background.
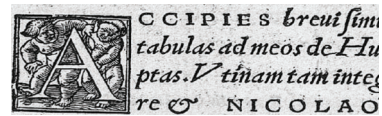

Fig. 1. Example of a lettrine.

By analyzing lettrines, the indexing of the digitized historical books of the beginning of the printing period can be ensured. Another interest of investigating lettrines in historical document images is to enrich semantically them by adding meta-data or semantic annotations. Thus, lettrines can be described, classified and compared using the obtained signatures, and historical documents can subsequently be dated historically, authenticated or characterized by identifying the differences between the analyzed lettrines. Other uses include developing relevant content-based image retrieval (CBIR) systems for lettrines. The idea consists in providing a lettrine image query to the developed lettrine CBIR system that will retrieve within a database all similar lettrines. There are several other needs expressed by the research historian community. For instance, by analyzing lettrines, the font, color or alphabet that characterizes the printer can also be deduced and investigated. In other cases, analyzing lettrines allows grouping the alphabet used in lettrines, studying the wear of buffers used to print lettrines, investigating the drawing painted in lettrine background, examining the progress of the used printing techniques, *etc.*

Nevertheless, the analysis of lettrines is considered complex, since there is a large variety and wide range of lettrine models and styles, and they contain a lot of information (e.g. texture, letter and decorated background). Coustaty *et al.* [4] considered the task of the extraction of the letter from a lettrine as not a straightforward one, due to the large variability of the representation of the information: colors, shapes, connectivity to different illustrations, *etc.* In the context of the MADONNE[1] and NaviDoMass[2] research projects, a

---

[1]http://madonne.univ-lr.fr
[2]http://navidomass.univ-lr.fr

number of studies have been carried out to extract specific parts from these complex graphic images and to compute signatures for indexing historical graphical images [5], [4], [6].

For instance, Coustaty *et al.* [4] presented a method for the extraction of drop cap letters by decomposing the information contained in the analyzed lettrines into several layers (*i.e.* segmenting the letter and the elements from its background) to characterize them by using a relevant signature. Their method was composed of three main steps: *(i)* simplification of the images using layer decomposition techniques, *(ii)* extraction of shapes from one of these layers, and *(iii)* selection of the shape that corresponds to the character to be extracted. It was based on a Meyer decomposition and Zipf modeling. The Meyer decomposition was used to filter out the noise, to extract the spatial frequencies of lettrine images and to segment them into separate layers (shape layer, texture layer and noise layer). Then, the use of Zipf law on the gray-levels of the shape layer ensured the detection of large homogeneous regions, which contain the letter. However, the main disadvantage of their method is that finding the most suitable encoding process when using Zipf law, which is capable of distinguishing different kinds of lettrines, is not a straightforward step (*i.e.* need of several parameters to be set in advance). In addition, it is quite difficult to analyze the resulting curve from Zipf law in order to decompose the shape layer of lettrines. Moreover, it requires a training task, and it is based on the definition of complex criteria and rules (e.g. the letter is a large connected component (CC), which is centered in the involved lettrine, and it does not touch the lettrine borders).

Another interest of examining lettrines has recently been revealed by historian exerts, which consists in the computer-aided discrimination of lettrines according to their style for lettrine indexing. For instance, Coustaty and Ogier [6] combined the use of the Zipf law with the bag of patterns. By extracting a set of the most relevant patterns (with pre-defined size masks) from lettrine images and ordering them according to the TF-IDF frequency criteria, they proposed a supervised method to retrieve lettrines according to their ornamental background style. Even if they obtained 80% of mean recall rate with a good precision, their method required a training task applied to a relevant set of lettrine images for pattern retrieval. Moreover, they used a quantization step to reduce the number of possible patterns.

Extracting a detailed description of lettrine characteristics suited to the experts' needs has become a major issue for lettrine categorization. Thus, there is a raising interest in the development of tools for lettrine clustering or categorization. For instance, Jouili *et al.* [5] proposed a structural-based framework, based on a graph-matching task. They proposed a graph-based representation for lettrines and compared their proposed representation with a statistical representation (*i.e.* graph) based on the generic Fourier descriptor (GFD). They concluded that the results provided from the use of the GFD are the lower ones and the structural-based representation is more appropriate to handle images of lettrines than the statistical one. Based on the developed method by Coustaty *et al.* [7] for letter extraction, Jouili *et al.* [5] used the most significant shapes, which were deduced from the six extracted largest CCs to define the graph node/vertex. Then, each vertex was described by a 4-D feature vector, representing the topolog-

ical features (e.g. coordinates of the center of mass, area, width/height ratio of each shape). Subsequently, the obtained graphs, characterizing the involved lettrines, were compared using a graph dissimilarity for lettrine clustering according to their ornamental background style. Even if their method was more effective than that proposed by Riesen and Bunke [8], their main drawback is the non-use of the topological relationships between the extracted shapes (*i.e.* graph vertices), which had a negative impact on the efficiency of their method.

The focus of this work is to propose generic solutions for lettrine indexing and particularly for recognition and classification. As a consequence, a novel fast bottom-up segmentation method, based on texture, is proposed for representative homogeneous region extraction from lettrines. The proposed method aims to extract the letter in a lettrine and the elements of its ornamental background for lettrine indexing. Based on the extracted lettrine components, two applications can be implemented for managing effectively a corpus or collections of lettrines. First, the separation of the letter from the elements of the ornamental background in a lettrine can be achieved. Secondly, a structural representation can be defined for characterizing lettrines. This structural representation is based on filtering automatically relevant information by extracting representative homogeneous regions from a lettrine, to generate a graph-based signature. The proposed signature provides a rich and holistic description of the lettrine structure and content by integrating varying low-level features (e.g. texture). Then, to categorize and classify lettrines with similar style, structure (*i.e.* ornamental background) and content (*i.e.* letter), a graph-matching paradigm has been carried out to compare and classify the resulting graph-based signatures according to the letter and the style of the background.

The remainder of this article is organized as follows. Section II describes the proposed method for representative homogeneous region extraction from lettrines. In Section III, based on the proposed method for representative homogeneous region extraction from lettrines, two deduced applications are presented for lettrine recognition and unsupervised lettrine classification. Section IV details firstly the experimental protocol. Then, qualitative results and an assessment of the two applications of the proposed method are presented. Subsequently, an analysis of the obtained results is discussed. Finally, our conclusions and future work are presented in Section V.

II. PROPOSED METHOD

Historian experts state that recognizing the letter and the style of the ornamental background of lettrines is an important step to retrieve similar lettrines and to identify the printer and the period. Therefore, the letter and the style of the ornamental background of lettrines are considered as the most important elements to analyze and investigate. Hence, it is obviously interesting to propose a generic and automatic method for letter extraction to meet the needs of historian experts.

Nevertheless, the use of a CC analysis technique for a simple letter extraction is not an appropriate choice due to the presence of noise and degradation on lettrines, their unclear boundaries, their complexity and the existence of a large variety and a wide range of lettrine models, *etc.* Moreover, based on the historian experts' findings, lettrines can be decomposed by two main components: *(i)* the ornamental background part,

which is characterized by regular frequencies, *(ii)* the letter and the other information part, which are characterized by low-level spatial frequencies and low gray-level variation. As a matter of fact, it is more appropriate to use a combination of a texture analysis technique for investigating the background part and a CC analysis technique for examining the letter part. Therefore, we propose a bottom-up segmentation method based on texture and CC analysis for representative homogeneous region extraction from lettrines.

A. Pre-processing and auto-correlation feature extraction (Step 1)

First, a lettrine is fed as input and is read as a gray-scale image. The extraction of texture information is processed on gray-scale lettrine images without introducing a binarization task. A binarization step is avoided because it causes a loss of information specifically the textural information. Then, to deal with pixels at image borders when computing the auto-correlation features on the whole image, a border replication step is introduced. Five auto-correlation features are extracted (e.g. main orientation of the rose of directions, intensity of the auto-correlation function for the main orientation, variance of the intensities of the rose of directions, mean stroke width and height along specific directions), which have been detailed in our previous work [9]. The auto-correlation descriptors are extracted from the pixels of the gray-scale lettrine images at four different sizes of sliding windows in order to adopt a multi-scale approach. Therefore, a 20-D feature vector (5 auto-correlation indices × 4 sliding window sizes), which corresponds to the results of the auto-correlation attribute extraction, is associated to every pixel from the lettrine image.

B. Pixel clustering and labeling (Step 2)

The obtained auto-correlation vectors for each lettrine image in the *Step* 1 are afterwards normalized and used in an unsupervised classification approach by using the hierarchical agglomerative clustering (HAC) algorithm and by setting the maximum number of clusters equal to 2. The idea consists in integrating an unsupervised task that ensures the pixel labeling for segmenting lettrine pixels in foreground and background ones (*cf.* Figures 2(f), 2(g), 2(h), 2(i) and 2(j)). Our objective is to retrieve the foreground pixels which illustrate the letter and pattern (*i.e.* ornamental background). The foreground pixels have been labeled in green color in Figures 2(f), 2(g), 2(h), 2(i) and 2(j).

C. Post-processing (Step 3)

Since the pixel clustering and labeling step (*Step* 2) has been performed, our output data consists of a pixel-labeled lettrine image. Nevertheless, we aim in this step to identify few groups of pixels sharing similar textural properties to extract homogeneous regions. Thus, a task of extracting and analyzing the CCs with an adaptive run-length smoothing algorithm (ARLSA) are carried out to fill automatically the space within each CC and subsequently to define regions (*cf.* Figures 2(k), 2(l), 2(m), 2(n) and 2(o)).

D. Representative homogeneous region extraction (Step 4)

The idea of this step starts by selecting representative foreground CCs in order to keep only a sub-set of the initially extracted CCs in the *Step* 3. By performing the majority voting

technique on each extracted CC, based on the majority label among the CCs pixels (obtained in the *Step* 2), a label is firstly assigned to each extracted CC. If the assigned label to an involved CC is similar with that one representing the largest CC, and the involved CC has at least 5% of the number of pixels of the largest CC, the involved CC is retrieved. This step aims at keeping the relevant information by selecting representative foreground CCs. This relevant information can represent the letter and representative regions describing the ornamental background. Therefore, by extracting the relevant foreground CCs, representative homogeneous regions are determined and labeled (*cf.* Figures 2(p), 2(q), 2(r), 2(s) and 2(t)).
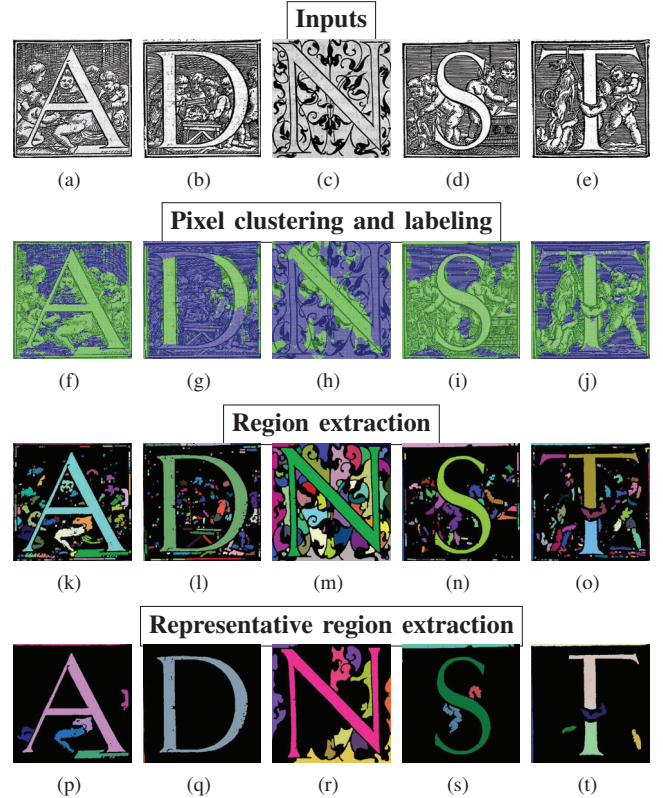

Fig. 2. Illustration of the intermediate resulting lettrine images derived from the proposed method for lettrine recognition and classification.

III. LETTRINE RECOGNITION AND CLASSIFICATION

Based on the set of representative homogeneous regions extracted in the *Step* 4 and by analyzing subsequently the extracted homogeneous regions, two following applications are targeted.

A. Lettrine recognition

Many historians have considered that it is interesting to retrieve lettrines in terms of letter criterion. Hence, it is important to extract efficiently and automatically the letter from a lettrine for lettrine recognition. According to Coustaty *et al.* [4], the largest extracted CC can cover a significant portion of a letter or a face in a lettrine. As a consequence, the largest extracted foreground region is retrieved based on the set of representative homogeneous regions extracted in the *Step* 4.

B. Unsupervised lettrine classification

Leveraging on the numerous advantages offered by using a structural representation instead of a statistical or ontology one [10], [5], a structural signature is proposed in this work. The proposed signature is used to represent a lettrine image with the set of representative homogeneous regions extracted in the *Step* 4. It is in the form of a directed attributed graph. The graph vertices correspond to the extracted representative homogeneous regions. Each vertex is described by a 66-D feature vector, representing the varying low-level features:

- 20 auto-correlation attributes to characterize the graphical characteristics of the extracted representative homogeneous region by computing the average of pixel auto-correlation indices extracted in the *Step* 1 [9],

- 46 shape, geometric and topological attributes to describe the shape and spatial properties of the extracted representative homogeneous region (topological position of the centroid in the x- and y-axis, number of pixels, contour area and perimeter, gray-level average and standard deviation, topological position of the bounding rectangle of the pixel set of the extracted region in the x- and y-axis, height, width and area of the bounding rectangle of the pixel set of the extracted region, ratio of the height to width of the bounding rectangle, ratio of the height (resp. width) of the bounding rectangle of the pixel set of the extracted region to the height (resp. width) of the analyzed lettrine image, 10 spatial moments, 7 central moments, 7 central normalized moments, 7 Hu moments).

Furthermore, a set of edges is built based on the topological relationships connecting the extracted representative homogeneous regions. An edge is built between two vertices, if the interaction existence and level (F) between two extracted representative homogeneous regions is significant. There is no interaction between two small regions unless they are close to each other, however a large region can have multiple interactions with more distant regions. Indeed, F models the gravitational force and implements the previous idea. To characterize the topological relationships between two extracted regions, few topological edge attributes are computed, F and absolute differences between the two extracted region centroids in the x- and y-axis.

To categorize and group lettrine images with similar style, (*i.e.* ornamental background) and content (*i.e.* letter), graph-based lettrine signatures are generated for the involved lettrine images and afterwards compared using a graph dissimilarity. In our experiments, we use the graph edit distance (GED). The GED is used to measure the (dis)similarity between the obtained graph-based lettrine signatures [11]. The GED deals with the computation of the minimum-cost sequence of the elementary graph editing operations (e.g. insertion, deletion and substitution of vertices or edges) to transform a graph to another one. Therefore, the evaluation of the proposed lettrine signature has been carried out based on firstly computing a distance matrix, whose elements represent the dissimilarity between the compared graphs. Our goal is to separate the involved lettrine images into k clusters. K is the number of lettrine classes defined by the research historian experts. It represents the different styles according to the ornamental

background appearances and properties of the involved lettrines.

IV. EVALUATION AND RESULTS

To demonstrate the robustness of the proposed solutions for lettrine recognition and classification, and provide additional insights into their accuracies, an experimental evaluation has been conducted using a dataset of 301 lettrine images. This dataset is collected from the *Centre d'Etudes Supérieur de la Renaissance* of Tours, France[3].

Firstly, the first application of the proposed method of homogeneous region extraction has been evaluated based on categorizing our experimental corpus according to the alphabet characters, since our first goal consists in letter based retrieval for lettrine recognition. Hence, this dataset is composed of 18 different alphabet characters.

Secondly, the lettrine images from this dataset have been labeled by research historian experts to define an appropriate ground truth. The defined ground truth aims to evaluate the second application of the proposed method of homogeneous region extraction, which consists in unsupervised lettrine classification, and to illustrate the effectiveness of the proposed graph-based lettrine signature. It has been set according to the number of classes provided by the historian experts. The number of classes provided by historians is equal to 3. This categorization is based on characterizing lettrines by the different styles according to their ornamental background appearances and properties. The three main lettrine styles, used in our experiments, are illustrated in Figure 3. Hence, our experimental corpus is composed of:

- 191 lettrine images having the first style "Style 1" (*cf.* Figure 3(a)),

- 95 lettrine images having the second style "Style 2" (*cf.* Figure 3(b)),

- 15 lettrine images having the third style "Style 3" (*cf.* Figure 3(c)).

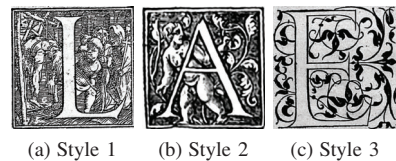(a) Style 1   (b) Style 2   (c) Style 3

Fig. 3.   Illustration of few examples of lettrines for each style in our experimental corpus.

A. Lettrine recognition

The evaluation of the first application of the proposed method of homogeneous region extraction from lettrines relies on using an open-source OCR system (Tesseract) at the resulting segmented lettrine image of the letter extraction task (*cf.* Section III-A and Figures 4(f), 4(g), 4(h), 4(i) and 4(j)). As a consequence, the character recognition accuracy (*RA*) rate is computed as an evaluation criterion of the performance of the proposed method for lettrine recognition. By comparing our results with those obtained with a state-of-the art method proposed by Coustaty *et al.* [4], a 17.6% gain of *RA* is noted, when

applying an OCR system for lettrine recognition (*cf.* Table I). A 84.7% mean of *RA* is obtained. We deduce that the proposed method gives better results for letter extraction application than the state-of-the-art method [4]. A visual comparison of the results using the proposed method for letter extraction and the one proposed by Coustaty *et al.* [4] indicates that better results are obtained with our method (*cf.* Figure 4). On the other hand, even the results show a much greater discriminating power for separating a letter from a lettrine using our method comparing a state-of-the-art method, both methods can not extract the "T" letter from the input lettrine (*cf.* Figure 2(e)), leading to confusion with the "r" letter when using an OCR system (*cf.* Figures 4(j) and 4(e)). This may be explained by the high complexity of the lettrine under consideration (*i.e.* overlapping boundaries of the pattern in the ornamental background and letter).
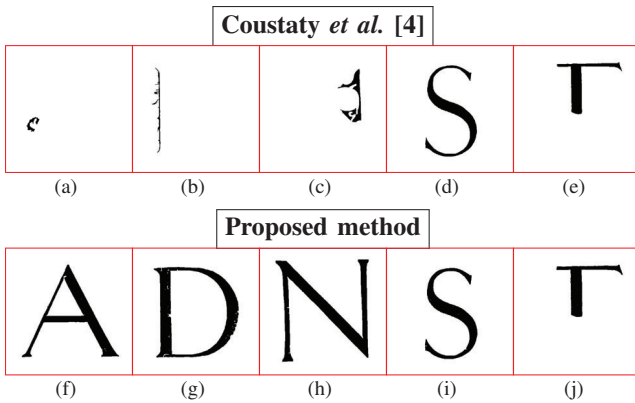

Fig. 4. Illustration of the resulting lettrine images derived from the proposed method of homogeneous region extraction for lettrine recognition.

TABLE I. EVALUATION OF THE PROPOSED METHOD OF HOMOGENEOUS REGION EXTRACTION FOR LETTRINE RECOGNITION.

	Evaluated methods	Tesseract (%)
Character recognition	Coustaty *et al.* [4]	67.1
accuracy (*RA*)	Proposed method	84.7

B. Unsupervised lettrine classification

To get an insight into the classification accuracy of the second application of the proposed method of homogeneous region extraction, which consists in unsupervised lettrine classification, and to illustrate the effectiveness of the proposed graph-based lettrine signature (*cf.* Section III-B), a confusion matrix is computed (*cf.* Table II). The confusion matrix illustrates the distribution of classes provided by the research historians for unsupervised lettrine classification. The following classification accuracy measures are computed: precision (P), recall (R), F-measure (F) and classification accuracy (CA). P_i and R_j denote the individual cluster precision and recall, respectively. The overall result is quite encouraging, since we obtain 65%(P), 82%(R), 73%(F) and 72%(CA). Nevertheless, a lower precision and a higher recall for the "Style 1" class are obtained, while a lower recall and a higher precision for the "Style 2" class are shown. Thus, we show that the proposed application tends to confuse between these two classes. This may be explained by the high similarity of the ornamental background appearance of lettrine belonging to the two classes: "Style 1" (*cf.* Figure 3(a)) and "Style 2" (*cf.* Figure 3(b)). Nevertheless, by comparing our results with those obtained with a supervised method proposed by Coustaty *et al.* [6], we

obtain promising accuracy results with a gain in computation time when using our method based on graph-based signature. A higher recall has been observed with the unsupervised classification approach used for the clustering of the obtained graph-based signatures.

TABLE II. EVALUATION OF THE PROPOSED GRAPH-BASED SIGNATURE FOR LETTRINE CLASSIFICATION.

		Ground truth			
		Style 1	Style 2	Style 3	
Clustering outcomes	Cluster 1	124	67	0	$\leftrightarrow P_1 = 0.65$
	Cluster 2	9	86	0	$\leftrightarrow P_2 = 0.91$
	Cluster 3	3	6	6	$\leftrightarrow P_2 = 0.4$
		$\updownarrow$	$\updownarrow$	$\updownarrow$	
		$R_1 = 0.91$	$R_2 = 0.54$	$R_3 = 1$	

V. CONCLUSIONS AND FURTHER WORK

In this paper, the issues relating to the analysis of lettrines is considered. As a consequence, we propose a bottom-up segmentation method, based on analyzing texture features, for the identification of representative homogeneous regions or similar groups of pixels sharing some visual characteristics from lettrines. Based on the extracted representative homogeneous regions, the separation of the letter from the elements of the background in an ornamental letter has been firstly carried out. Secondly, a structural representation is proposed for characterizing a lettrine to generate a graph-based signature. The proposed signature provides a rich and holistic description of the lettrine style by integrating varying low-level features (e.g. texture). Then, by comparing the different graph-based signatures, the lettrines with similar style are grouped. The experimental results have shown that the proposed bottom-up lettrine segmentation method provides a better performance than the state-of-the-art methods for lettrine recognition and unsupervised lettrine classification. Further work needs to be done to assess the proposed solutions with other texture descriptors. In addition, we will evaluate the proposed solutions for lettrine recognition and classification on a wider style variety of lettrines.

REFERENCES

[1] N. Journet, J. Ramel, R. Mullot, and V. Eglin, "Document image characterization using a multiresolution analysis of the texture: application to old documents," *IJDAR*, pp. 9–18, 2008.

[2] V. Eglin, S. Bres, and C. Rivero, "Hermite and Gabor transforms for noise reduction and handwriting classification in ancient manuscripts," *IJDAR*, pp. 101–122, 2007.

[3] S. Uttama, P. Loonis, M. Delalandre, and J. M. Ogier, "Segmentation and retrieval of ancient graphic documents," in *GREC*, 2006, pp. 88–98.

[4] M. Coustaty, R. Pareti, N. Vincent, and J. M. Ogier, "Towards historical document indexing: extraction of drop cap letters," *IJDAR*, pp. 243–254, 2011.

[5] S. Jouili, M. Coustaty, S. Tabbone, and J. M. Ogier, "NaviDoMass: structural-based approaches towards handling historical documents," in *ICPR*, 2010, pp. 946–949.

[6] M. Coustaty and J. M. Ogier, "Discrimination of old document images using their style," in *ICDAR*, 2011, pp. 389–393.

[7] M. Coustaty, J. M. Ogier, and R. P. N. Vincent, "Drop caps decomposition for indexing: a new letter extraction method," in *ICDAR*, 2009, pp. 476–480.

[8] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *IVC*, pp. 950–959, 2009.

[9] M. Mehri, P. Gomez-Krämer, P. Héroux, and R. Mullot, "Old document image segmentation using the autocorrelation function and multiresolution analysis," in *DRR*. SPIE, 2013.

[10] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *IJPRAI*, pp. 265–298, 2004.

[11] H. Bunke and K. Riesen, "Towards the unification of structural and statistical pattern recognition," *PRL*, pp. 811–825, 2012.