Local Enlacement Histograms for Historical Drop Caps Style Recognition

Michaël Clément, Mickaël Coustaty, Camille Kurtz, Laurent Wendling

Local Enlacement Histograms for Historical Drop Caps Style Recognition

Michaël Clément, Mickaël Coustaty, Camille Kurtz, Laurent Wendling

Local Enlacement Histograms for Historical Drop Caps Style Recognition

Michaël Clément ; Mickaël Coustaty ; Camille Kurtz ; Laurent Wendling

***Abstract*—This article focuses on the specific issue of drop caps image recognition in the context of cultural heritage preservation. Due to their heterogeneity and their weakly structured properties, these historical images represent challenging data. An important aspect in the recognition process of drop caps is their background styles, which can be considered as discriminative features to identify both the printer and the period. Most existing methods for style recognition are based on low-level features such as color or texture properties. In this article, we present a novel framework for the recognition of drop caps style based on features of higher levels. We propose to capture the spatial structure carried by these images using relative position descriptors modeling the enlacement between local cells of pixel layers obtained from a document segmentation step. Such descriptors are then exploited in an efficient bag-of-features learning procedure. Experimental results obtained on a dataset of historical drop caps images highlight the interest of this approach, and in particular the benefit of considering spatial information.**

I. Introduction

The cultural heritage of Europe is a unique public good that represents our collective memory, and constitutes a solid basis for the development of the industries relying on digital contents. In this paper, we deal with the problem of cultural heritage preservation, focusing on collections of heterogeneous and weakly structured documents. Such complex data raise both the information research issue and the navigation problem among these corpus. More specifically, this work focuses on historical images of ornamental letters, called *drop caps* or *lettrines*. As illustrated on the first row of Fig. 1, drop caps images are usually made up of the following main characteristics: the letter, the color of the letter and the background patterns around the letter. In general, such background patterns can be categorized into different classes, which are illustrated on the second row of Fig. 1: *hatchings*, *dotted*, and *decorative* patterns (with either black or white letters). An important step in the indexing process of drop caps consists in annotating these different background styles, as they are used by historians to retrieve similar looking drop caps, to identify the printer and the period.

In this article, we propose a new method to describe and classify historical drop caps images according to their background styles (independently from the letter). To this end, the rest of this article is organized as follows. In section II, we give an overview of related works and of our contributions. The first step of our approach consists in a segmentation strategy based on Zipf law to decompose drop caps into three layers of information in the images. This procedure is presented in section III. Then, we introduce in section IV how to describe such layers using local enlacement histograms and following a bags-of-features strategy. Experimental results attesting of the performance of this approach will be found in section V. Conclusion and perspectives are given in section VI.

Fig. 1: Illustrative examples of drop caps (first row) and of different background styles for the same letter (second row).

II. Related Work

The automated analysis of graphical drop caps is a promising step towards the digitalization of historical books and manuscripts. Indeed, such information can be used to date historically, to authenticate, and to characterize books by identifying the differences between analyzed historical drop caps. In order to enrich drop caps semantically, by adding meta-data or semantic annotations, many works proposed to describe, to classify and to compare them using some statistical or structural signatures. This context therefore encompasses the ongoing development of content-based image retrieval (CBIR) systems for historical drop caps [1]. The development of such image retrieval systems is complex, since these historical drop caps present a large variety and wide range of models and styles, and because the images contain a lot of information (e.g. texture, letter, and decorated background).

In order to cope with the complexity of these images, the authors of [2], [3] proposed an original binarization method based on connected operators which enable to extract relevant document objects by means of the component-tree structure. This work allowed for an efficient shape-based classification of the image connected components, focusing in particular on detecting and recognizing the letters. However, the main drawback of this approach is the necessity to choose the color of foreground or background, while drop caps could have either a white or black letter on a white or black background.

To deal with this limitation, Coustaty *et al.* [4] proposed a method for the decomposition of drop cap letters by extracting the information contained in the images into several layers (i.e. segmenting the letter and the elements from its background).

The Meyer decomposition allows used to filter out the noise, to extract the spatial frequencies of drop cap images, and to segment them into separate layers (shape, texture, and noise layers). Then, the use of Zipf law on the most frequent gray level patterns of the shape layer ensured the detection of large homogeneous regions, which often contain the letter.

In parallel of this document analysis context, bags-of-features strategies have attracted numerous research attentions in the computer vision community for object recognition and image classification tasks [5]. The approach was notably adapted to historical drop caps images, to recognize styles of strokes [6] as well as letters and background styles in [7], [8], in combination with graph-based representations.

In the field of symbol recognition, the works of [9], [10] introduced bags-of-relations (BoR), an original way to produce vocabularies of spatial relations. The approach was applied on a well-controlled set of visual primitives specific to the application domain (*e.g.*, circles, corners or extremities of symbols). A generalization of the bags-of-relations approach was proposed in [11] using Force Histograms [12]. These works illustrated the interest of considering spatial relations descriptors into the bags-of-features framework. A new relative position descriptor was recently proposed in [13], and illustrated notably the interest of considering enlacement and interlacement configurations to classify the style of drop caps images. However, the descriptors were applied globally on the whole images, which does not allow to capture distinctive patterns depicted in drop caps at a more local scale.

In the present work, we propose to further explore this idea of exploiting spatial relations to recognize the background styles of historical drop caps. Our contributions are to consider enlacement histograms across local cells (instead of evaluating the images globally) in combination with the Zipf decomposition strategy, and to exploit this characteristic spatial information into a bags-of-features framework.

III. Drop Caps Decomposition

As stated before, graphical drop caps are composed of three main layers of information which need to be extracted while their content can be white or black. Based on the results presented in [4], we decided to use a Zipf law segmentation process in order to extract three layer of information and to analyze the spatial configuration. In this section, we present the employed approach to decompose drop cap images into meaningful layers of pixels representing respectively homogeneous regions, outline and details (H, O and D).

Zipf law [14] is an empirical law relying on a power law. It relies on the idea that in phenomena figured by a set of topologically organized symbols, the distribution of the occurrence numbers of n-tuples named patterns is organized in such a way that the frequencies of the patterns M_1, $M_2 ... M_n$, noted N_1, $N_2 ... N_n$, are in relation with the rank of these symbols when sorted with respect to their occurrence frequencies. The following relation holds:

$$N_\sigma(i) = k \times i^a \quad (1)$$

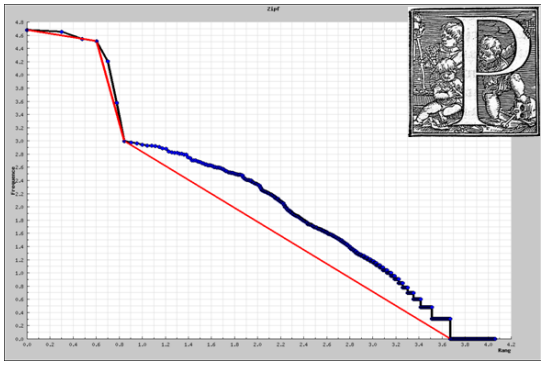


Fig. 2: Example of a drop cap image along its Zipf plot. In blue are indicated the different 3×3 patterns sorted by appearance frequencies in the image. In red is the corresponding Zipf law, which can be separated into 3 linear parts.

where the number of occurrences of the pattern with rank i is represented by $N_\sigma(i)$, k is a constant linked to the length of the symbol sequence studied, and a a constant that characterizes the value of the exponent. Even if the relation is not linear, a simple transformation step leads to a linear relation between the logarithm of N_σ and the logarithm of the rank. Finally, the value of the exponent a can be easily estimated by the leading coefficient of the regression line approximating the experimental points of the 2D graph $(\log_{10}(i), \log_{10}(N_\sigma(i)))$, with i varying from 1 to n.

In our context of drop caps image, the patterns considered are all the possible gray level patterns of size 3×3 which can appear in an image. As the images we are dealing with in this study are mostly binary, we decided to keep only 3 gray levels. This step is essential to build a realistic Zipf curve associated with a drop cap (as by keeping all the 256 gray levels available would lead to patterns with very low frequencies without any meaning). More details about this step can be found in [4]. The obtained graph, presented in Fig. 2 is called Zipf graph. The approximation of the graph is made using the least square method on the experimental points. As a matter of fact, we can see that the curve cannot be reasonably modeled by a straight line. Nevertheless, we can observe that three different linear segments can be extracted to model the curve. Then, according to the frequency of the pattern, we can distinguish three sets of patterns for which the Zipf law holds.

In Fig. 3, we show for different drop caps images the associated decompositions into three layers according to the Zipf curve. We can find an interpretation to these layers. The first set (black layer) which is associated with the left part of the curve, corresponds to the most frequent patterns. It usually corresponds to flat regions with homogeneous intensity values. The second set (red layer), associated with the middle part of the curve, seem to correspond to outline regions, where a lot of transitions between dark and light pixels can be found in the image. Finally, the right part of the curve (green layer), which corresponds to the third layer, can be interpreted as details of an image, *i.e.* the least frequent patterns. As a result, for each drop cap image we obtain a decomposition into three

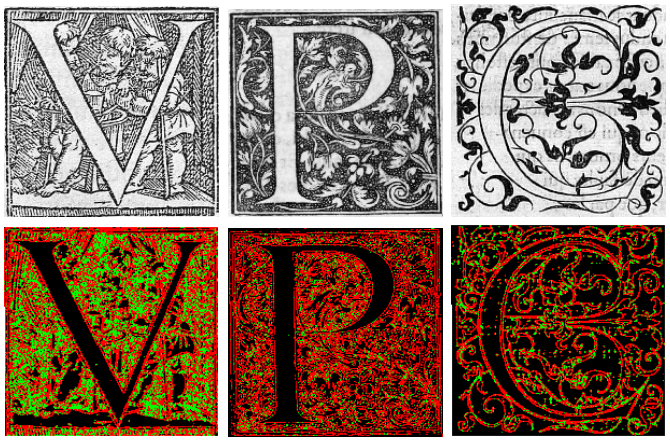


Fig. 3: Examples of Zipf decompositions of drop cap images. The first row shows drop caps from the different background styles (from left to right: *hatchings*, *dotted* and *decorative*). The second row shows their corresponding Zipf decompositions into three layers (in black: *homogeneous* layer H; in red: *outline* layer O; in green: *details* layer D).

meaningful pixel layers: the *homogeneous* layer H, the *outline* layer O and the *details* layer D.

In the following, we propose to keep all of these three levels of information, and to encode the pairwise spatial organization between them to describe the content of drop caps images. In particular, this allows to retain small details such as very thin contours or small dots, which are characteristic of certain background styles.

IV. Local Enlacement Learning

In this section, we propose a way to describe the complex content of drop caps images by considering the spatial organization of layers of pixels at a local scale. For this purpose, we first present enlacement histograms, which are spatial relations descriptors allowing to characterize how two objects are imbricated in each other. Then, we show how to incorporate such descriptors into a bags-of-features framework to recognize drop caps styles.

A. Enlacement histograms

We briefly present the model used to describe the relative enlacement of objects, which was initially introduced in [13].

Two-dimensional objects of the plane are considered, where an object A is defined by its characteristic function $f_A : \mathbb{R}^2 \to \mathbb{R}$. This generic definition allows to handle complex objects composed of multiple connected components (for instance in the binary case, an object can be a set of pixels). We represent the oriented line of angle θ at the altitude ρ by the non-finite set $\Delta^{(\theta,\rho)} = \{e^{i\theta}(t + i\rho), t \in \mathbb{R}\}$. The subset $A \cap \Delta^{(\theta,\rho)}$ represents a one-dimensional slice of the object A, also called a *longitudinal cut*. In the case of crisp objects, such a longitudinal cut of A is either empty (the oriented line does not cross the object) or composed of a finite number of segments. In the general case of real-valued objects, a longitudinal cut of A along the line $\Delta^{(\theta,\rho)}$ is defined as:

$$\begin{array}{rccl} f_A^{(\theta,\rho)} : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & t & \longmapsto & f_A(e^{i\theta}(t + i\rho)). \end{array} \quad (2)$$

The goal is to describe how an object A is enlaced by another object B. The idea is to capture the occurrences of points of A being *between* points of B. To determine such occurrences, the objects are handled in a one-dimensional case, using longitudinal cuts. For a given oriented line $\Delta^{(\theta,\rho)}$, we seek to combine the quantity of object A (represented by $f_A^{(\theta,\rho)}$) located simultaneously *before* and *after* object B (represented by $f_B^{(\theta,\rho)}$). Let f and g be two bounded measurable functions with compact support from $\mathbb{R}$ to $\mathbb{R}$. The enlacement of f with regards to g is defined as:

$$E(f, g) = \int_{-\infty}^{+\infty} g(x) \int_{x}^{+\infty} f(y) \int_{y}^{+\infty} g(z)\, \mathrm{d}z\, \mathrm{d}y\, \mathrm{d}x. \quad (3)$$

The scalar value $E(f_A^{(\theta,\rho)}, f_B^{(\theta,\rho)})$ represents to which degree A is enlaced by B along the oriented line $\Delta^{(\theta,\rho)}$. For crisp objects (*i.e.*, each point is either 0 or 1), this corresponds to the total number of ordered triplets of points $(b_i, a_k, b_j), i < k < j$, which can be seen as arguments to put in favor of the proposition *"A is enlaced by B"* in direction θ. Algorithmically, this value can be derived by an appropriate distribution of segments lengths along the longitudinal cuts of both objects (see [13] for more details).

To further measure the global enlacement of two objects in direction θ, we aggregate the one-dimensional enlacement values obtained for all parallel lines $\{\Delta^{(\theta,\rho)}, \rho \in \mathbb{R}\}$. The enlacement of A by B in direction θ is therefore defined by:

$$\mathcal{E}_{AB}(\theta) = \frac{1}{\|A\|_1 \|B\|_1} \int_{-\infty}^{+\infty} E(f_A^{(\theta,\rho)}, f_B^{(\theta,\rho)})\, \mathrm{d}\rho, \quad (4)$$

where $\|A\|_1$ and $\|B\|_1$ denote the respective areas of the objects. This normalization factor allows to achieve invariance with regards to scaling transformations. Then, the computation of $\mathcal{E}_{AB}$ along a set of k discrete directions in $[0, \pi]$ yields a circular histogram describing the enlacement of A by B. Note that both histograms $\mathcal{E}_{AB}$ and $\mathcal{E}_{BA}$ must be considered to fully describe the spatial configuration.

B. Learning by Bags-of-Enlacement

We present how to extend the bags-of-relations approach to work with the Zipf segmentation results, using enlacement histograms as features. The approach is inspired by traditional bags-of-features strategies usually applied with local image descriptors (such as SIFT or HOG).

A given training image is first segmented using the Zipf decomposition approach which was presented previously. We therefore obtain three binary layers of the image, representing respectively *homogeneous* regions, *outline* and *details* (H, O and D). Then, the images corresponding to the three possible couples of layers are considered, denoted by HO, HD and OD in the following. Each of these images is split into local, non-overlapping square cells of size 32×32. For a drop cap image

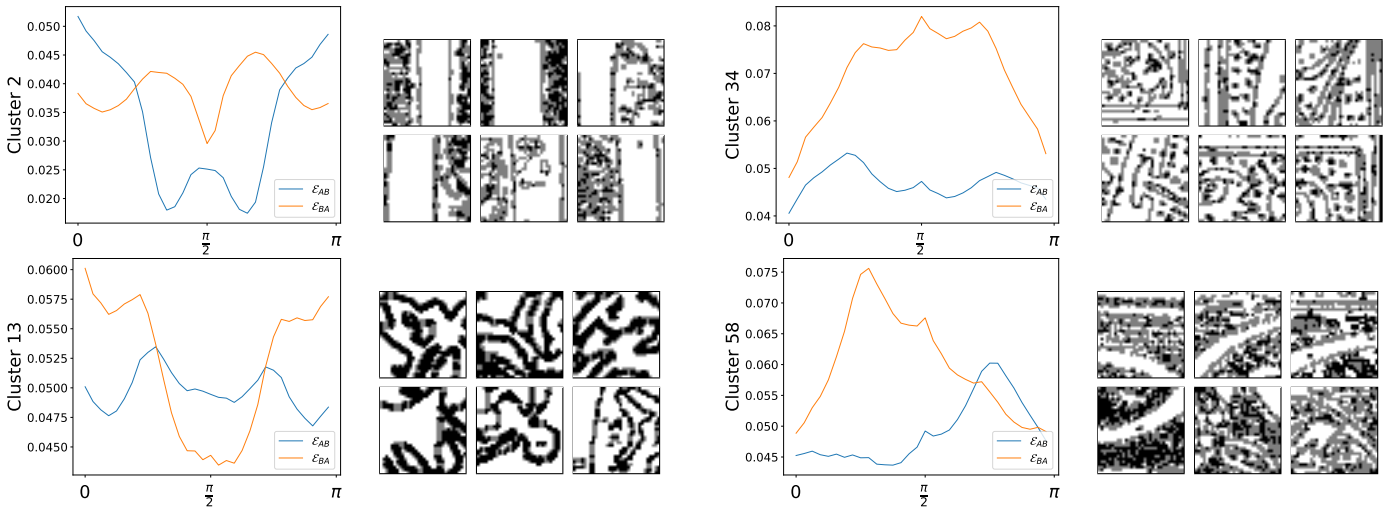


Fig. 4: Illustrative examples of local enlacement configurations obtained for the HO layers (homogeneous regions in white and outline in gray). For each of the four illustrative clusters, the centroid of the cluster is shown, alongside with some corresponding local cells that are attached to this centroid. Clustering was performed for a vocabulary size of $K = 100$, and with $\mathcal{E}_{AB}$ and $\mathcal{E}_{BA}$ concatenated into a single descriptor (shown here on separate curves for the sake of visualization).

of size 200×200, this results in about $36 \times 3 = 108$ cells (*i.e.*, 36 blocks for each of the three couples of layers). For each of these cells, we compute the two enlacement histograms $\mathcal{E}_{AB}$ and $\mathcal{E}_{BA}$ between the two binary objects represented in the current layer (*i.e.*, for a local cell of HO for example, A and B correspond to homogeneous and outline pixels respectively). As these two histograms are complementary to describe the spatial configuration, they are concatenated into a single feature vector describing the local enlacement between the two layers of a given cell.

We then apply the K-Means clustering algorithm [15] to regroup enlacement descriptors into similar clusters, therefore building a spatial vocabulary of local enlacement configurations. We perform distinct clusterings of enlacement histograms, one for each couple of layers. That is, for a vocabulary of size K, we perform three independent clusterings with the descriptors coming from couples of layers HO, HD and OD respectively. We therefore produce a total of $K \times 3$ vocabulary words corresponding to similar local enlacement configurations between the different layers of the Zipf decomposition. For visualization purposes, some examples of local enlacement words obtained for the HO layers (homogeneous regions and outline) are presented in Fig. 4. We can notice how very distinctive drop caps patterns are captured by these spatial vocabulary words.

Following this vocabulary building procedure, we then apply a pooling step which consists in representing each image by a distribution of each vocabulary word it is composed of. In this work, we applied a soft encoding strategy [16] that allows to take into account the distance of the descriptors to the built clusters. For a given image, the soft assignment value for the vocabulary work $\mathbf{w}_k$ is given by:

$$z_k = \frac{1}{n} \sum_{i=1}^{n} K_\sigma(\mathbf{f}_i, \mathbf{w}_k) \tag{5}$$

where $\{\mathbf{f}_i\}_{i=1}^n$ denote the set of descriptors of the image (*i.e.*, its set of enlacement histograms across the local cells), and where K_σ is the Gaussian kernel:

$$K_\sigma(\mathbf{x}_1, \mathbf{x}_2) = \exp - \frac{||\mathbf{x_1} - \mathbf{x_2}||^2}{2\sigma^2}, \tag{6}$$

with σ allowing to control the smoothness of assignments. The image is finally described by the feature vector $\mathbf{z} = [z_1, \ldots, z_K]$ summarizing the soft assignment values for the K vocabulary words. As for vocabulary building, this encoding procedure is followed independently for each couple of layers HO, HD and OD. We therefore obtain three codebooks corresponding different layers of information in the images. Finally, style recognition of test images can be performed by using any supervised classifier trained on a given codebook, or on a combination of multiple ones (by concatenating the feature vectors of the different codebooks).

V. Experimental Validations

In this section, we report experimental results obtained for recognition of drop cap styles with our approach, and we compare these results with relevant baselines.

A. Dataset

The dataset used in the following experiments is a set of 636 images of historical drop caps. These images were classified by historians into three classes, each class representing a different background styles (independently from the letter): *hatchings* (253 images), *dotted* (214 images) and *decorative* (169 images). The decorative class is composed of various

TABLE I: Average classification accuracy scores obtained for the recognition of drop caps styles, for different combinations of layers issued from the Zipf decomposition, and for different vocabulary sizes.

Layers / K	100	200	500
DH	79.81 $\pm$ 2.15	80.04 $\pm$ 1.63	80.00 $\pm$ 1.55
OD	80.69 $\pm$ 1.82	80.96 $\pm$ 1.88	81.30 $\pm$ 1.62
OH	80.96 $\pm$ 1.83	81.03 $\pm$ 1.68	81.43 $\pm$ 1.64
DH+OH	81.49 $\pm$ 1.87	81.70 $\pm$ 1.62	81.43 $\pm$ 1.38
OH+OD	82.68 $\pm$ 1.62	82.68 $\pm$ 1.65	82.49 $\pm$ 1.99
DH+OD	84.84 $\pm$ 1.35	83.98 $\pm$ 2.39	84.53 $\pm$ 1.56
DH+OH+OD	85.07 $\pm$ 2.05	84.32 $\pm$ 2.33	85.26 $\pm$ 1.10

patterns, with notably several inversed images (*i.e.*, dark letter and outline over light background, and conversely). To maintain homogeneity in the dataset, all images were downsampled to a maximum size of 200×200.

B. Experimental Protocol

First, the images of the dataset are segmented following the Zipf decomposition strategy presented in section III. For a given image, we obtain three layers of decomposition (H, O and D) which are combined into the couples of layers HO, HD and OD, and split into local cells of size 32×32. Then, we compute the enlacement histograms for each of these cells. All enlacement histograms are computed onto a set of 32 discrete directions, equally separated along the $[0, \pi]$ interval. When concatenating $\mathcal{E}_{AB}$ and $\mathcal{E}_{BA}$, this translates into descriptors of size 64. Following this indexing phase, we obtain a total of 21503 descriptors over the entire dataset.

The dataset is then split into a training set representing 25% of its total size (159 training images), while the remaining 75% is used for testing. The bags-of-enlacement (BoE) approach is applied on training images for all possible combinations of couples of layers, and for different vocabulary sizes (*i.e.*, the value of K for K-Means clustering). For soft encoding (see Eq. 5 and 6), we used an adaptive Gaussian kernel $K_{\tilde{\sigma}}$ where $\tilde{\sigma}$ was fixed according to the mean of the pairwise Euclidean distances between the training enlacement features and their associated vocabulary words.

The resulting codebooks are then fed to SVM classifiers [17] to recognize the background styles of the corresponding drop caps images. In these experiments, we used linear SVMs following a *one-versus-all* strategy for multiclass classification. The soft-margin hyperparameter of SVMs was optimized by performing grid search on the training set, following a 5-fold cross validation. Different SVM kernels were tried (Gaussian and χ^2) but they did not seem to improve the classification results, while greatly increasing the computation times.

To better assess the robustness of the approach, this classification procedure is repeated 10 times with different randomly chosen training sets (each time preserving the distribution of the different classes). We evaluate the classification results using the accuracy score for each run, and we report the average scores over all runs.

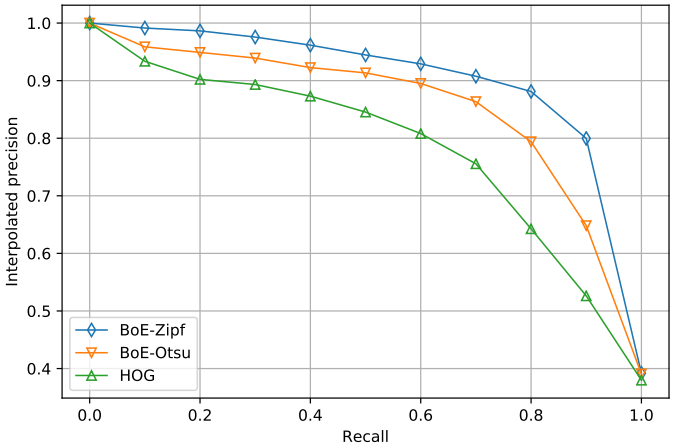


Fig. 5: Interpolated precision-recall curves obtained for our method (BoE and Zipf) and for the considered comparative baselines (Otsu thresholding instead of Zipf, and standard HOG bags-of-features).

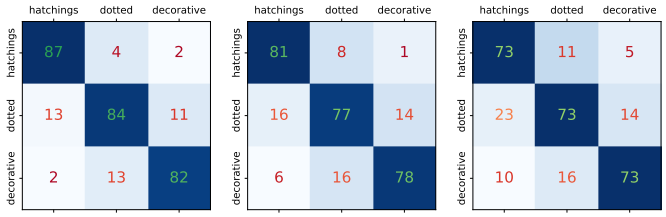
C. Experimental Results

Table I reports the average classification accuracy scores obtained for all the possible combinations of layers from the Zipf decomposition, and for different vocabulary sizes in the bags-of-enlacement approach. The first three lines correspond to the individual couples of layers DH, OH and OD, each of them being composed of independent vocabularies of K words. The following lines denote combinations of these couples of layers, obtained by concatenating their respective codebooks before SVM training (and therefore increasing the dimensionality of the representations).

From these results, we can observe the overall stability of the accuracy scores across the 10 random training runs, attesting of the generalization capacity of the proposed learning approach. Altogether, the recognition results seem to increase as we combine more couples of layers, the best results being obtained for the full combination DH+OH+OD. This suggests that all three layers of information can be useful to characterize the different background patterns represented in drop caps images. Regarding the vocabulary sizes, we can also observe that the recognition results remain particularly stable across the tested values of 100, 200 and 500. This suggests that most spatial information is contained in a small number of vocabulary words.

D. Comparative Study

As our approach is composed of two main parts (Zipf decomposition for segmentation, and bags-of-enlacement for recognition), we propose baseline comparisons on these two aspects. For Zipf decomposition, we compare the results to the same bags-of-enlacement approach but applied directly on Otsu binarizations of drop caps images. For bags-of-enlacement, we compare our results to classical HOG features [18] extracted from the base gray-level drop cap images. The same non-overlapping cells of size 32×32 that were used



(a) BoE-Zipf (b) BoE-Otsu (c) HOG

Fig. 6: Average confusion matrices obtained for our proposed method and for the considered comparative baselines.

for local enlacement descriptors are considered. These features are then pooled into a similar bags-of-features framework with soft encoding. In the following comparative results, BoE-Zipf indicates our proposed approach (bags-of-enlacement over the Zipf decomposition) with an initial vocabulary size of $K = 100$, and with all couples of layers combined (DH+OH+OD), therefore comprising a total of 300 spatial words. For comparability purposes, BoE-Otsu and HOG comparisons are therefore performed with $K = 300$ vocabulary words, and following the same classification protocol.

Fig. 5 shows the precision-recall curves obtained for the proposed comparative baselines. In the context of this drop caps style recognition application, these results confirm the interest of considering local enlacement descriptors based on the spatial configuration of pixel layers, instead of gradient-based features which are classically used in object recognition tasks (*i.e.*, BoE vs HOG). We can also observe that the Zipf decomposition seem to perform slightly better than Otsu thresholding of images. This suggests that this method allows to better extract the complex patterns depicted in these historical drop caps. To better grasp the relative performance of these comparative methods, Fig. 6 also shows the respective confusion matrices, allowing to evaluate the performance on a class-by-class basis. We can remark that our approach seem to recognize *hatchings* relatively well, which are indeed directional alternating patterns that are efficiently characterized by enlacement histograms.

VI. Conclusion

In this article, we proposed an approach for the recognition of historical drop caps images, focusing in particular on their background styles, independently from the letters. To this end, we used a segmentation strategy based on Zipf law to extract meaningful pixel layers in the images. These layers of pixels are described across local cells using relative position descriptors called enlacement histograms. Such descriptors are then learned with a bags-of-features strategy in order to classify the images. Experimental validations on a dataset of drop caps images showed the interest of (1) using the Zipf decomposition strategy as opposed to a standard binarization; and (2) considering enlacement histograms as local descriptors instead of standard gradient-based features.

In future works, we plan to extend this approach by increasing the size and the variety of this drop caps dataset. In collaboration with historians, the goal would be to characterize a more diverse range of background patterns according to historical preservation needs. Another perspective of this research would be to study how to refine the produced spatial vocabularies, for instance by putting an emphasis on most relevant words, and by labeling them with semantic annotations.

References

[1] S. Jouili, M. Coustaty, S. Tabbone, and J.-M. Ogier, "NAVIDOMASS: structural-based approaches towards handling historical documents," in *International Conference on Pattern Recognition (ICPR)*, pp. 946–949, 2010.

[2] B. Naegel and L. Wendling, "Combining Shape Descriptors and Component-tree for Recognition of Ancient Graphical Drop Caps," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, pp. 297–302, 2009.

[3] B. Naegel and L. Wendling, "A Document binarization method based on connected operators," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1251–1259, 2010.

[4] M. Coustaty, R. Pareti, N. Vincent, and J.-M. Ogier, "Towards historical document indexing: Extraction of drop cap letters," *International Journal on Document Analysis and Recognition*, vol. 14, no. 3, pp. 243–254, 2011.

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[6] T. T. H. Nguyen, M. Coustaty, and J.-M. Ogier, "Bags of Strokes Based Approach for Classification and Indexing of Drop Caps," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 349–353, 2011.

[7] M. Coustaty and J.-M. Ogier, "Graph matching versus bag of graph: A comparative study for lettrines recognition," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 356–360, 2015.

[8] M. Mehri, P. Gomez-Krämer, P. Héroux, M. Coustaty, J. Lerouge, and R. Mullot, "A bottom-up method using texture features and a graph-based representation for lettrine recognition and classification," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 226–230, 2015.

[9] K. Santosh, B. Lamiroy, and L. Wendling, "Integrating vocabulary clustering with spatial relations for symbol recognition," *International Journal on Document Analysis and Recognition*, vol. 17, no. 1, pp. 61–78, 2014.

[10] K. Santosh, L. Wendling, and B. Lamiroy, "BoR: Bag-of-Relations for Symbol Retrieval," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 28, no. 6, 2014.

[11] M. Clément, C. Kurtz, and L. Wendling, "Bags of Spatial Relations and Shapes Features for Structural Object Description," in *International Conference on Pattern Recognition (ICPR)*, 2016.

[12] P. Matsakis and L. Wendling, "A new way to represent the relative position between areal objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 634–643.

[13] M. Clément, A. Poulenard, C. Kurtz, and L. Wendling, "Directional Enlacement Histograms for the Description of Complex Spatial Configurations between Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017 (in press).

[14] G. K. Zipf, *Human Behavior and the Principle of Least Effort*. Hafner Pub. Co, 1949.

[15] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Berkeley Symposium on Mathematical Statistics and Probability (BSMSP)*, pp. 281–297, 1967.

[16] J. C. V. Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek, "Visual Word Ambiguity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1271–1283, 2010.

[17] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, 2005.