



HAL
open science

Bags of Strokes Based Approach for Classification and Indexing of Drop Caps

Thi Thuong Huyen Nguyen, Mickaël Coustaty, Jean-Marc Ogier

► **To cite this version:**

Thi Thuong Huyen Nguyen, Mickaël Coustaty, Jean-Marc Ogier. Bags of Strokes Based Approach for Classification and Indexing of Drop Caps. ICDAR 2011, Sep 2011, Beijing, China. pp.349-353. hal-00715953

HAL Id: hal-00715953

<https://hal.science/hal-00715953v1>

Submitted on 9 Jul 2012

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

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

Bags of Strokes Based Approach for Classification and Indexing of Drop Caps

Thi Thuong Huyen Nguyen Mickaël Coustaty Jean-Marc Ogier

L3i Laboratory

Avenue Michel Crepeau

17042 La Rochelle, France

Email: {thi_thuong_huyen.nguyen, mickael.coustaty, jean-marc.ogier}@univ-lr.fr

Abstract—This paper proposes an approach to process drop cap images – images of decorated letter that begin chapters of old documents that are preserved in libraries, museums – in the domain of characterization, classification and indexing of old documents. The originality of our proposal is based on the fact that we do not try to extract the letter of drop caps but to classify the drop caps according to period, author and style. The drop caps are characterized by using relevant visual features such as length, thickness, orientation, complexity and change of direction on their primitive elements: strokes. The purpose of this approach is to efficiently extract information embedded in the drop caps for the classification and the indexing of old documents. These new visual features based on bags of strokes are more easily calculable and generally applicable than texture or shape features. Experiments based on characterization, classification and indexing phases demonstrate the performance of our propositions and the advances that they represent in terms of content-based drop caps retrieval.

Keywords—old documents; drop caps; characterization, classification and indexing; codewords; feature detection

I. INTRODUCTION

With the expansion of computers and powerful storage devices, requirement to converse and to exploit thousands of pages of documents, especially old documents printed for a long time, becomes indispensable and urgent. Despite their small size, the drop caps play a very important role in the work of historians because they change depending on time of printing, style of authors or printers. So, if relevant visual characteristics of drop caps are extracted, one can navigate easily in a large mass of documents by using only these drop caps. In reality, even if there are many image search engines that rely on visual contents of images such as color, texture, shape or spatial information [1], we cannot apply directly methods for processing normal images on the drop caps due to their specific properties. Texture features (e.g. cooccurrence features [2] and Wavelet based features [3]) and shape features (e.g. Fourier descriptors [4] and moment invariants [5]) have been applied widely in content-based image retrieval systems. However, texture features are effective only for uniform texture images or homogeneous regions; while shape features are only effective for images with simple and clean object contours.

This paper addresses the analysis of drop caps by visual features. The original idea of this work is to consider that

basic information of drop caps is no more pixels but strokes. Each drop cap contains a group of strokes or black/white dots on a white/black background that are the result of a pen on paper or in this case, of an impression seal on paper (See Figure 1). Therefore, these strokes do not represent the illuminance information like texture and shape features of real-world images. Texture captures spatial repetition of same pattern in different directions in space whereas shape represents a specific feature that is related to object contour. Neither of them pays attention to specific information represented in the strokes of drop caps in general.



Figure 1. Drop Caps Examples

The importance in the recognition process of the drop caps in this article is not to segment separately the letter and the elements of the background but to classify and index the drop caps into categories without a priori information. In this article we limit our emphasis on the characterization of the drop cap images to simplify the step of classification and indexing by using our proposed visual features of the strokes. The rest of the paper is organized as follow. In section 2, background and motivation are described. Features extraction based on our method are presented in section 3. Section 4 gives the experimental results over 1000 drop caps images. Conclusions are drawn in section 5.

II. BACKGROUND AND MOTIVATION

In the context of the grand project NAVIDOMASS which is a collaboration between different French laboratories since 2007, we assessed more pieces of work in processing and indexing of drop caps and of old documents. Early researches focused on segmentation and classification of different types of ancient documents [6]. This work proposed to use a hybrid segmentation algorithm that separates a page into two parts: one of the shapes which describe connected components and the other corresponding background. From

the first stage of segmentation, the interaction with users is required to collect information from different types of images in old documents. Thus, the effectiveness and the availability of this approach depend on feedback of users and empirical results. More particularly, the work [7] presented a method allowing images of old documents based on texture analysis to be characterized. Its main idea is to retain characteristics of different areas at different levels of frequency and orientation. This approach does not require a priori hypotheses about the structure of analyzed pages. The work [8] has the same intention of indexing historical documents. However, instead of using the characteristics of frequency and orientation, they try to find significant areas within a page by decomposing it into simple elements and by extracting geometric attributes such as eccentricity. The constructed graph of these attributes is used to structure a spatial description of each image and to look for similar images. From the results of these previous works, there are many attempts to analyze and index historical documents that focus just on drop caps [9] by the characterization of the drop caps using primitives of texture (i.e. strokes) and not at the global level of texture analysis.

As we discussed earlier, since there is no effective approach of texture analysis at global level in solving the problem of drop caps classification and indexing, we proposed to work on their primitive strokes, based on the simple idea that our images are not classical, but composed of strokes, instead of being composed of pixels, like natural scene images. Strokes have some key features (e.g. length, thickness, curvature) and they can be assimilated to edges of normal images after edge detection. To characterize edges, there are several methods in literature. The main idea of shape descriptors CSS (i.e. Curvature Scale Space) in [10] is to find an image CSS that consists of several arch-shape contours representing the inflection points of the shape as it is smoothed. The maxima of these contours are used also to represent a shape with its other features like area, circularity, eccentricity, major axis orientation. Another method, described in [11], extracts structural features of a shape in spectral domain by using different shape signatures such as complex coordinates, curvature function or centroid distance to produce standardized Fourier coefficients. In addition, Zernike moments given by [12] are based on the theory of orthogonal polynomials whose principle is similar to the Fourier transform, that is to deploy the signal into a series of orthogonal basis. While these approaches can capture important features of shape, there are always some difficulties when we apply them to the drop caps processing. This is due to the fact that strokes are not always closed and often contain different branches. These shape descriptors are therefore more adequate to the shape of simple objects than to the strokes in drop caps. The article [13] shows a new attempt to solve the problem of complex edges in images by using a set of features called edge/structural

features whose principle lies between texture and shape. Examples of edge/structural features include but not limited to: edge length, edge complexity, loop in the edges and their statistics throughout the image. Using the water-filling algorithm, these features are extracted directly without the representation of the closed shape. This idea inspires us to improve the characterization of strokes in the drop caps by applying some their key features like length, complexity, etc.

The drop cap images in old documents are often complex and damaged over time. Thus, the first step in drop caps processing (i.e. restoration and noise removal) is necessary. According to [14], decomposition of image in geometric and oscillatory components for finding significant components is a key objective in the drop caps processing. From this idea, [15] used an approach developed by [16] to separate original drop caps into three layers of information that are easier to process (i.e. Shape Layer, Texture Layer and Highly Oscillating Layer). The Texture layer helps to highlight texture of drop caps and emphasizes many areas of strokes or patterns. Thus, this work is carried out on that layer because of its importance in the characterization of drop caps.

III. ALGORITHM TO EXTRACT STROKE SIGNATURES AND TO CLASSIFY DROP CAPS

The preprocessing of images is a preliminary step before extracting the relevant features and it includes several auxiliary operations. Specifically, the main operations to be performed are binarization, noise reduction and skeletonization. The binarization permits to take off less important parts like noise, uniform background in the images and to emphasize sudden changes of gray level like strokes. The noise reduction is carried out by selecting the size of connected components that are regarded as noises and by removing at once in background areas and in strokes of the binary images. Since the thickness of strokes is variable, the step of skeletonization becomes necessary. This process gives us information in details on the size and on the shape of each stroke. Furthermore, by using the Euclidean distance transform [17], we also obtain the skeleton of strokes, which is one of the major objects of the later step of characterization, and of course the characteristics of stroke thickness [18].

There are many features that we can calculate from the obtained skeleton (e.g. length, thickness, curvature, circularity, eccentricity, major axis orientation). However, if we take too much features for each skeleton, we may encounter the problem of redundant features and this brings about the drop caps recognition problem later. After many tests, we have chosen four main characteristics (i.e. length, thickness, major axis orientation and complexity) and a polynomial approximation to describe the change in direction and to characterize the global peculiarities of the skeleton.

- **The length** of a skeleton is calculated by the straight-line distance between pixels (the Euclidean distance)

and by noting the difference of distance between pixels in the 8-connected neighborhood area (i.e. 8-neighbors) [19] (see Figure 2(d)).

- **The complexity** of a skeleton is measured by the ratio of the number of branching points on the skeleton and its length : $C = \frac{X}{L}$ such as C – the complexity of the skeleton; X – the number of branching points of the skeleton; L – the length of the skeleton.
- **The average thickness** of the skeleton is easily obtained from the distance map issuing from the Euclidean distance transform [17].
- According to [22], we can calculate **the major axis orientation** of the skeleton using the angle α formed by the first axis of inertia of the skeleton with the horizontal axis (see Figure 2(a)). The first axis of inertia can be found by a Principal Component Analysis (PCA).
- To calculate the **change of direction** of the skeleton, the most common method is the Freeman code. It describes the change of direction at pixel level and then it is strongly attached to the number of pixels of each skeleton. The comparison of the change in direction of different skeletons requires coding standardization even using the absolute or relative Freeman code. Moreover, the absolute Freeman code is suddenly changed because of the rotation of the skeleton. Thus, the polynomial approximation for the Freeman chain is a good choice to characterize the change in direction of the skeleton. Freeman code is not influenced by the rotation and the polynomial approximation permits to keep a common size for all the skeletons. There exists only a small problem referring to the structure of multiple branches on the skeleton (See Figure 2(c)). To solve this problem, we consider that if there are several branches on the skeleton, we find the most dominant branch which is closer to its major axis orientation (See Figure 2(b)), and then we calculate the change in direction by the Freeman code and its polynomial approximation.

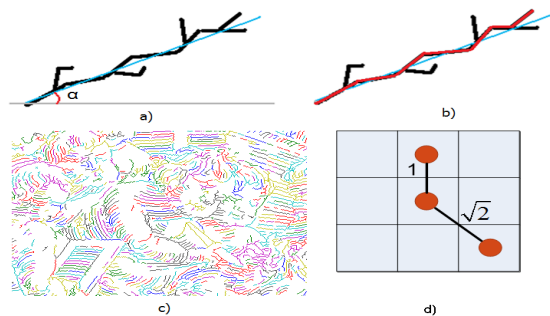


Figure 2. Characteristics of skeletons – (a) Major axis orientation of a skeleton and the angle formed by this axis with the horizontal axis – (b) The dominant branch of a skeleton – (c) Complicated structure of multiple branches of skeleton – (d) Euclidean distance between two pixels

After calculating all the features of the skeleton, a normalization step is necessary in order to compare all the skeletons. The obtained values of all four characteristics (i.e. length, thickness, major orientation and complexity) are limited in the interval $[0, 1]$. Similar descriptors are computed for all skeletons of all images in the database.

There is still an important step to facilitate the drop caps retrieval and to estimate the difference between the drop caps of each period of history, each author and each style. Our idea of this approach to recognize drop caps comes from the efficiency of text retrieval systems [26], [24]. It consists in using the automatic classification of strokes characteristics to group similar drop caps. From the set of descriptors extracted from all the skeletons of the image, the classification of feature vectors is carried out. Our goal is to choose groups of relevant feature vectors that represent the different types of skeletons and that are considered as "codewords" of each image [20]. We have here the notion of "bags of strokes". The aim of this first classification is to reduce number of feature vectors of an image and computation time for the second feature vectors classification of all images. After grouping the characteristic vectors of all the skeletons, we choose a representative of each group by calculating the average value of its elements. These representatives are regarded as the feature vectors of each image. Then the second classification plays an important role in grouping representatives together to build a dictionary of representative "codewords" of all drop caps. If we combine these two steps of classification in one step, computation time for vectors classification is exceeded while the number of "codewords" stays the same. We use the same algorithm for non-supervised sequential classification [21], [23] to accomplish these two steps of classification. Based on the dictionary of "codewords" of all drop caps, each image is then represented by a vector with components given by the frequency of occurrence of the "codewords" that the image contains. A query image is treated in the same way. Its vector of "codeword" frequencies is computed. Matching images are obtained by measuring the similarity between the query and all images in the database vectors.

IV. EXPERIMENTAL RESULTS

For our experiments, we use an image data set of 1000 drop caps with a ground-truth of 12 classes achieved by hands based on kind of strokes, density and style of decoration in each drop cap to evaluate the results of retrieval. This ground-truth presents the disadvantage to be constructed in a subjective manner, and there are therefore poor estimates for different drop caps. In addition, the evaluation of our experiments is quite difficult because of the heterogeneity of user requirements such as the demand for retrieval results of historian researchers which is quite different than normal user's requirements. However, we consist in using these experiments to estimate the general performance of the

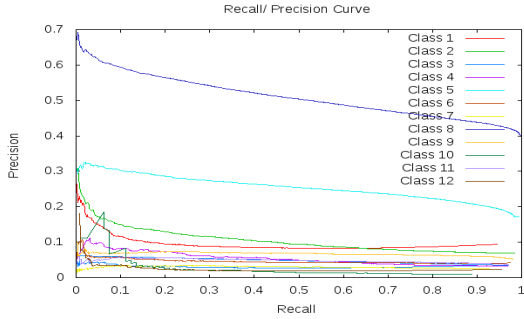


Figure 3. Precision/Recall curve of 12 classes at a low threshold of classification step

retrieval system and the effectiveness of our system will be improved later by the feedback of historians researchers and of users. In the context of the NAVIDOMASS project and for a system of drop caps retrieval, we evaluate our performance by using two widely used metrics, i.e. Precision and Recall of a recognition system and Precision/Recall graph [25].

At the first step of our evaluation, we try to estimate the performance of drop caps retrieval from the view of users by assuming that one user sees only the top 50 responses. Depending on used thresholds at the classification step, a summary table of maximum and minimum values of the precision of responses for all 12 classes is constructed as one can see in table I. If we examine all the responses of system, by using different thresholds of classification step (18 different thresholds for the classification step, their value range from minimum to maximum dissimilarity of all descriptors), we obtain many Precision/Recall curves presenting performance of our drop caps retrieval system (e.g. a Precision/Recall curve in Figure 3 at a low threshold for the classification step). From these curves, we can estimate the efficiency of each threshold of classification step and determine the most appropriate threshold to optimize performance of our system for the next step to get users feedback.

Table I
SUMMARY RESULTS OF PRECISION VALUES FOR ALL 12 CLASSES WITH THE FIRST 50 RESPONSES

Class	1	2	3	4	5	6
Max Value %	39.9	15.4	17.0	10.7	40.8	34.9
Min Value %	14.8	3.2	1.2	4.0	8.0	4.8
Amount of Images	87	157	39	22	368	49
Class	7	8	9	10	11	12
Max Value %	12.1	63.1	12.3	4.9	8.7	23.3
Min Value %	1.6	1.2	0.2	0.0	1.5	1.6
Amount of Images	9	33	22	65	32	31

For the next evaluation, we estimate influence of change of orientation or scaling of a query on the effectiveness of

retrieval. We will see if the drop cap query is changed, how it changes the result of retrieval with the original image and its modified images as one can see in Figure 4. Concerning the rotation, after we turn the drop cap queries, these rotated images are processed in the same way as other drop caps, and we obtain a Precision/Recall curve which is compared to that of the original drop cap (e.g. see Figure 5). We find that our methods provides good results compared to initial responses of the rotated image (green line), since the curve of the original image and the rotated one is almost identical. This means that the rotation does not influence strongly on the outcome of the retrieval on in any case, our algorithm is stable for rotation. For the scaling, to verify its influence on the result, we tried to change size of original image along two horizontal and vertical axes. From this graph in Figure 6, we find that if the original image is reduced in the same way along two axes, the result does not change much as two curves are almost the same. However, we cannot guarantee that if the reduced image is too small, the effectiveness remains the same because in this case, we lose a lot of stroke information. On the other side, for the change of scale along a single axis, the result changes completely. For some initial result, the performance is very high but in general, the retrieval efficiency is low. The main reason is that two general features of strokes (i.e. length and thickness) are completely changed. Moreover, the choice of scaling along one of the horizontal or vertical axis does not decide the outcome.

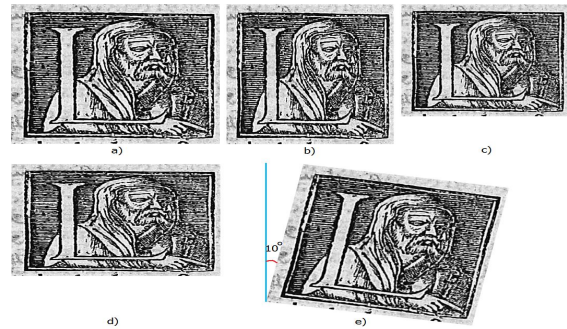


Figure 4. Request Images — (a) Original Image — (b) 80% scale of original image along axis X — (c) 80% scale of original image along two axis X and Y — (d) 80% scale of original image along axis Y — (e) 10 degrees rotation of original image (we limited rotation angle to 10 degree because it represents errors that can appear in scanning process)

V. CONCLUSION

This paper presents our proposal for a new processing chain of characterization, classification and indexing of drop caps without any a priori knowledge. The originality of our proposal is based on the fact that we do not try to extract the letter of drop caps but to classify the drop caps according to period of history, author and style. Thus, we describe how to characterize the visual contents of drop caps by basing

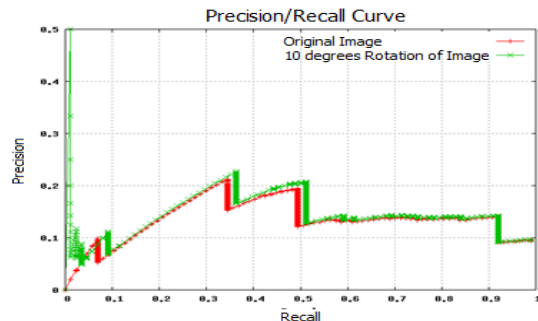


Figure 5. Precision/Recall curves of original image (red) and rotated image (green)

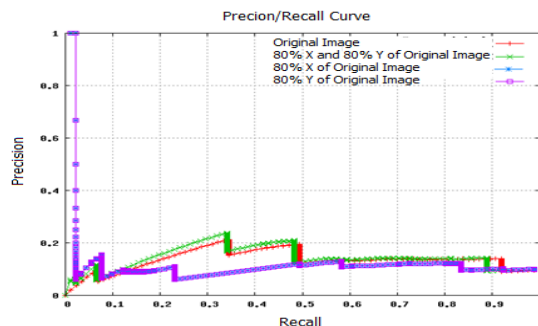


Figure 6. Precision/Recall curves of original image (red) and scaled images

on their primitive elements in an approach of codewords. Although we have not evaluated the general performance of our system, we try to estimate the effectiveness of each type of drop caps. From these estimates, we can find other methods to improve the performance of each type of drop caps in the retrieval.

ACKNOWLEDGMENT

This work was supported in part of NAVIDOMASS project by l'Agence Nationale de la Recherche Française (ANR-06-MDCA-012)

REFERENCES

- [1] M. Flickner et al., *Query by Image and Video Content: The QBIC System*, Computer, vol. 28, no. 9, pp. 23-32, Sept. 1995.
- [2] R.M. Haralick, K. Shanmugan and I. Dinstein, *Textural Features for Image Classification*, IEEE Transactions on System, Man and Cybernetics, vol. 3, pp. 610-621, 1973.
- [3] D. Charalampidis and T. Kasparis, *Wavelet-Based Rotational Invariant Roughness Features for Texture Classification and Segmentation*, IEEE Transactions on Image Processing, vol. 11, no. 8, Aug. 2002.
- [4] D. Zhang and G. Lu, *Shape-Based Image Retrieval using Generic Fourier Descriptor*, Signal Processing: Image Communication, vol. 17, no. 10, pp. 825-848, Nov. 2002.
- [5] M.K. Hu, *Visual Pattern Recognition by Moment Invariants*, IEEE Transactions on Information Theory, vol. 8, no. 2, pp. 179-187, Feb. 1962.
- [6] J.Y. Ramel, S. Leriche, M.L. Demonet and S. Busson, *User-Driven Page Layout Analysis of Historical Printed Books*, IJDAR, vol. 9, no. 2-4, pp. 243-261, 2007.
- [7] N. Journet et al., *Document Image Characterization Using a Multiresolution Analysis of the Texture: Application to Old Documents*, IJDAR, vol. 11, no. 1, pp. 9-18, 2008.
- [8] Salim Jouili, Mickael Coustaty, Salvatore Tabbone, and Jean-Marc Ogier, *NAVIDOMASS: Structural-based approaches towards handling historical documents*, ICPR 2010.
- [9] M. Coustaty, G. Nguyen et al., *Analyzing Old Documents Using a Complex Approach: application to lettrines indexing*, Advances in Knowledge Discovery and Management, to appear, 2011.
- [10] S. Abbasi, F. Mokhtarian and J. Kittler, *Curvature Scale Space Image in Shape Similarity Retrieval*, Multimedia Systems, vol. 7, no. 6, pp. 467-476, 1999.
- [11] D. Zhang and G. Lu, *A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-Based Image Retrieval*, JVCIR, vol. 14, no. 1, pp. 39-57, 2003.
- [12] M.R. Teague, *Image Analysis via the General Theory of Moments*, Journal of Optical Society of America, vol. 70, pp. 920-930, 1980.
- [13] X.S. Zhou and T.S. Huang, *Edge-Based Structural Features for Content-Based Image Retrieval*, Journal of Pattern Recognition Letters, vol. 22, no. 5, 2001.
- [14] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, American Mathematical Society Boston, 2001.
- [15] Coustaty, Mickael and Pareti, Rudolf and Vincent, Nicole and Ogier, Jean-Marc, *Towards historical document indexing: extraction of drop cap letters*, IJDAR, to appear, 2011.
- [16] S. Dubois et al., *Adding a Noise Component to a Color Decomposition Model for Improving Color Texture Extraction*, CGIV2008, pp. 394-398, Jun. 2008.
- [17] C.T. Huang and O.R. Mitchell, *A Euclidean Distance Transform Using Grayscale Morphology Decomposition*, IEEE TPAMI, vol. 16, no. 4, pp. 443-448, Apr. 1994.
- [18] A. McAndrew, *An Introduction to Digital Image Processing with MATLAB*, London : Thomson Learning, 2004.
- [19] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, New York, Massachusetts : Addison-Wesley, 1992.
- [20] L. Fei-Fei, R. Fergus and A. Torralba, *Recognizing and Learning Object Categories*, Short Course CVPR, 2007.
- [21] S. Theodoridis and K. Koutroumbas, *Pattern Recognition 3rd Edition*, Elsevier Academic Press, 2006.
- [22] S. Kanda et al., *Method for Determining Orientation of Contour Line Segment in Local Area and for Determining Straight Line and Corner*, US Patent 5586199, Fujitsu Limited (Kawasaki, JP), 1996.
- [23] E. Vi-Tong and P. Gaillard, *An Algorithm for Non-Supervised Sequential Classification of Signals*, Pattern Recognition Letters, vol. 5, no. 5, pp. 307-313, 1987.
- [24] J. Sivic and A. Zisserman, *Efficient Visual Search of Videos Cast as Text Retrieval*, IEEE TPAMI, vol. 31, no. 4, pp. 591-606, 2009.
- [25] C.D. Manning, P. Raghvan and H. Schtze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- [26] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, 1999.