HAL
open science

# Towards historical document indexing : extraction of drop cap letters

Mickaël Coustaty, Rudolf Pareti, Nicole Vincent, Jean-Marc Ogier

## To cite this version:

Mickaël Coustaty, Rudolf Pareti, Nicole Vincent, Jean-Marc Ogier. Towards historical document indexing : extraction of drop cap letters. International Journal on Document Analysis and Recognition, 2011, 14 (3), pp.243-254. 10.1007/s10032-011-0152-x . hal-00916007

**HAL Id: hal-00916007**

**https://hal.science/hal-00916007v1**

Submitted on 9 Dec 2013

**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.

**International Journal on Document Analysis and Recognition manuscript No.**
(will be inserted by the editor)

# Towards Historical Document Indexing: Extraction of Drop Cap letters

**Mickael Coustaty**[1] · **Rudolf Pareti**[2] · **Nicole Vincent**[2] · **Jean-Marc Ogier**[1]

Received: date / Accepted: date

**Abstract** This paper deals with the difficult problem of indexing ancient graphic images. It tackles the particular case of indexing dropcaps (also called Lettrines), and specifically considers the problem of letter extraction from this complex graphic images. Based on an analysis of the features of the images to be indexed, an original strategy is proposed. This approach relies on filtering the relevant information, on the basis of Meyer decomposition. Then, in order to accommodate the variability of representation of the information, a Zipf law modeling enables detection of the regions belonging to the letter, what allows it to be segmented. The overall process is evaluated using a relevant set of images, which shows the relevance of the approach.

**Keywords** Letter extraction · Historical documents

## 1 Introduction

There is a growing interest in digitally preserving and providing access to the historical document collections residing in libraries, museums, and archives. Such archives of old documents are a unique public asset, forming the collective and evolving memories of our societies. Indeed, ancient documents have a historical value not only for their physical appearance but also for their contents. Examples include unique manuscripts written by well-known scientists, artists or writers; letters, trade forms, or official documents that help to reconstruct historical sequences for a given place or time; and artistic items such as stamps, illustrations, covers, etc.

The challenge that is currently being addressed throughout Europe is the conversion of such heritages into digital libraries that enable them to be preserved, but also to make them available worldwide using web-based portals (like the impact project[1] for instance). Through the medium of better-designed digital libraries, citizens of the future should be able to gain access to a myriad forms of knowledge from anywhere, at any time, and in an efficient and user-friendly fashion. A number of initiatives exists focusing on the creation of large digital libraries that are globally accessible. Google is now engaged in a project to create a global virtual library. A number of European libraries have started a similar joint project [2]. DELOS is the European Network of Excellence for digital libraries [1]. The construction of such libraries has an additional important challenge: the analysis of documents and the extraction of knowledge. This goal involves projects to design and develop semantics-based systems to acquire, organize, share, and use the knowledge embedded in the documents. The field of Data Mining, combined with Document Analysis, offers a robust methodological basis for performing tasks such as descriptive modelling (clustering and segmentation), classification, discovering patterns and rules, and retrieval by content applied to document sources and databases. Old documents may be originals (paper, parchment etc.) or in image form (already scanned, possibly using now-outdated technologies). The key requirement is to be able to process these unique manuscripts, whether they are presented as free-flowing text (treatises, novels, ...) or structured at various levels of physico-logical structure correspondence (letters, census lists, trade forms, ...). Degradation may be caused by a lifetime of use, and access must

[1]Imedoc Team - L3i Laboratory
Avenue Michel Crepeau
17042 La Rochelle, France
E-mail: mcoustat, jmogier@univ-lr.fr
.
[2]SIP Team - LIPADE Laboratory
45, rue des Saints-Peres
75270 Paris Cedex 06, France
E-mail: nicole.vincent@mi.parisdescartes.fr

[1] http://www.impact-project.eu/

also be provided to user annotations and corrections, stamps and unique artwork. Each class of document requires a different approach throughout the conversion process and lends itself to different levels of information extraction and description. In summary, the work comprises the analysis of knowledge in historical documents to compile metadata that are then used to access digital libraries.

In the knowledge society, the objective is not simply to digitize documents but to create semantically-enriched digital libraries of such digitized documents. Enrichment of a document means the addition of semantic annotations to digital images of the scanned documents. Such metadata are intended to describe, classify and index documents by their content. It would then enable easy access to this cultural and scientific heritage from any place and at any time.

The main research goal of the French project "NaviDoMass" (NAVIgation into DOcument Masses) is thus to work in a collaborative framework on the Analysis of Old Documents. This goal consists of developing Pattern Recognition and Image Analysis techniques that allow the extraction of knowledge from documents and its conversion into Digital Libraries containing the scanned pages enriched with semantic information.

This paper deals with one specific point of the NaviDoMass Project, concerning the indexing of the graphic portions of historical documents. Even if we present our method on a specific corpus, one part of the strategy remains generic and could be applied to any other graphic images in an indexing scheme of historical documents. In this paper, we will use three terms to define the graphic objects of our corpus: *lettrine, drop cap and ornamental letter*. The literature contains examples of all three names being used for the same objects. The proposed approach is applied to lettrines, which are present in all books of the fifteenth and sixteenth century[2]. Analysing these images is thus significant for indexing books of the beginning of the printing period. The lettrines, which were originally used in a decorative goal, are actually used, on one hand, to distinguish the works, the printers and to detect the beginning of a chapter or a paragraph. On the other hand, historians aim at using lettrines to navigate into database from a lettrine to another, to make query by example or to analyse their specifications. These research activities are made in order to analyse the font, the color or the alphabet that characterize the printer.

Images of lettrines are made up of three principal elements: the letter, the pattern, and the background (see Fig. 1). An important step in the indexing of lettrines consists of segmenting the letter and the elements from its background, in order to characterize them by using a relevant signature.

[2] Historians of our projects plan to digitize 15000 books from this period. Actually, 479 books are available online at http://www.bvh.univ-tours.fr/. These books represent more than 168000 pages digitized

One approach may be to define different styles among the lettrines according to the background appearance. In fact, what would be a good criterion to achieve this clustering ? Historians criteria are not those that ease the computations. Color is not significant as, according to the lettrine, there can be no leading colors. It is possible to make clusters as done in some studies [9] but what is the best number of clusters ? This approach would induce ad hoc method for each lettrine family and on the whole the recognition rate of letter is not improved. Indeed, the greatest difficulty of our work is the degradation of the images and ideal patterns.

The focus of this paper is a consideration of the problem of the extraction of the letter from a lettrine (dropcap), which appears to be quite a simple problem but is in fact quite complicated, owing to the great variability of the representation of the information: colors, shapes, connectivity to various illustrations, etc. (see Fig. 1 for an illustration of various lettrines).

This paper presents in detail the several stages of our method:

1. Simplification of the images using layer decomposition techniques
2. Extraction of shapes from one of these layers
3. Selection of the shape that corresponds to the character.

This work is inspired by [20,21,10] who used a Wold decomposition on the one hand, and a Zipf law on the other hand, to extract the elements of drop caps. Our paper is organized as follows: first of all, we illustrate features of lettrines, in order to highlight the difficulty of extracting a letter. Based on these features, we present our strategy for the extraction of the letter, based on Meyer decomposition (layer segmentation), Zipf modeling , and letter extraction.

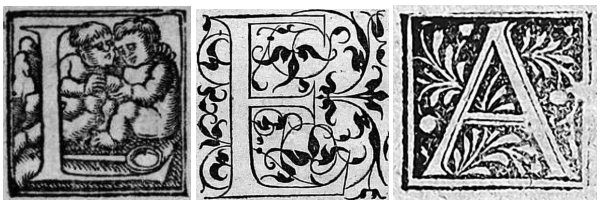


**Fig. 1** Various lettrines illustrating the great variability of representation of the information

## 2 Image description and features

Before considering our strategy, it is important to describe the images to be processed, and to understand their principal shared features, so that we may apply a generic method.

### 2.1 The support

The majority of our images were scanned at 300 dpi. The support for scanning is the original paper of the book, which may have become degraded over time. The paper is composed of three elements: the fiber, the filler, and the gluing (the filler and the gluing are added to give better properties to paper [22]). Fibers, mainly composed of organic materials such as linen, cotton or wood, are a cause of deterioration of the paper in ancient documents. Among the problems due to aging of these fibers, we may cite yellowing or thinning of the pages (which may cause a transparency of the paper and problems of over-print scanning) and the weakening of the pages. This weakening is often accompanied by mechanical effects such as tearing or creasing of the pages. The main consequences for images resulting from these degradations of the support, are (i) a very high level of noise, (ii) a possible over-printing of the verso page, and (iii) some possible soiled areas.

### 2.2 Printing technique

The second major feature of the documents that we are processing arises from the techniques used in early printing. At that time, stamps were hand-carved from wood to be inked and pressed on paper. These buffers, specific to each printer, are used today to authenticate documents and to characterize them. The buffers allowed documents to be created in black and white but did not allow levels of gray. In order to create shades and shadows on documents, and to suggest grays, printers used artistic strategies, replacing the shades of gray by parallel lines. For example, in the images in Figure 1 the shadows of the arms and legs of various characters are represented by parallel lines. This technique is used for different decoration elements present in the book, for example at the end of chapters or in the beginning of paragraphs.

### 2.3 Description of the Lettrines

Among various types of graphic images of old documents contained in [11], we are particularly interested in ***Lettrines***. They correspond to images widely used in books over time. A lettrine is an ornamental letter that begins a chapter or a paragraph (see Fig. 2) and can be viewed as a binary image composed of strokes. Some studies ([14,20,25,6]) have attempted to characterize this kind of graphic image. Discussions with our historian partners in the NaviDoMass project enable us to understand how historians decompose lettrines. According to these specialists, lettrines can be viewed as an overlaying of two layers: the background and the letter.

- The background may be black, white, dashed, or streaked, and is entirely composed of ornamental objects. It may be decorative or figurative. The ornamental objects may be considered themselves as set of layers ; at last, a global frame can be present around the image taking different styles. These different elements can be seen in figure 3 ;
- The letter is one of the most important pieces of a lettrine. It is homogeneous in colour (often black or white), and from a specific font. Most of the time, a simple extraction of connected component is not enough because the images are noisy, theirs boundaries are not clean and they are often split into many pieces.

Finally, these layers may be contained within a frame (see Fig. 3) that defines the boundaries of the typographical block. It can be composed of zero, one or two strokes.

From frequencial and spatial points of view, these layers can be separated. Background is composed of a mix of uniform and textured areas, while the letter correspond to an uniform white or black area in the middle of the lettrine.

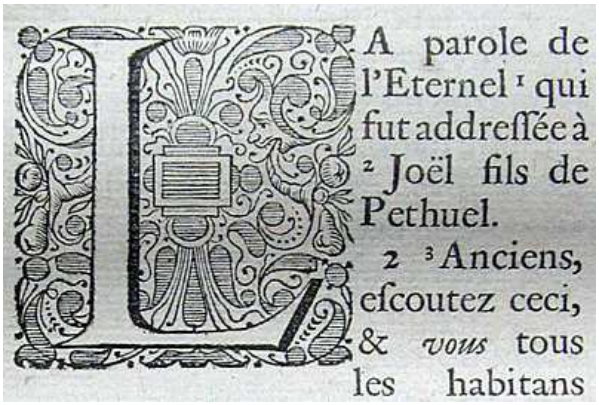
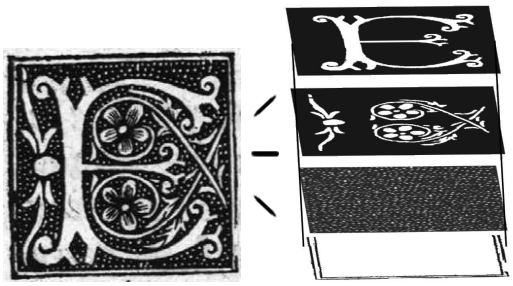



**Fig. 2** Example of lettrines in context



(a) Full Image (c) Separate layers

**Fig. 3** Layers in a lettrine. From bottom to top, we can observe the frame layer, the background layer, the pattern layer and the letter.

### 2.4 Important features for the definition of a relevant strategy

Considering features described above, one can establish the basis of a relevant strategy. First of all, considering the high probability of degradation of the useful information, a pre-processing technique should be applied, to filter the noise out of images. Pre-processing techniques could also be used for segmentation purposes, in order to separate the layers identified by the experts in this domain, *i.e.* historians. In this regard, examination of a sizeable set of lettrine images highlights the fact that the ornamental background part is characterized by quite regular frequencies (texture analysis), while the letter and all the other information correspond to very low-level spatial frequencies, and are characterized by quite large homogeneous areas.

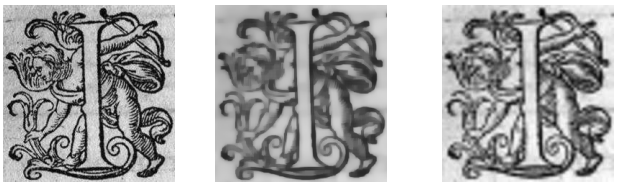
Another important feature that must be considered in letter extraction is related to the fact that the letter can be of different colors : black or white. Consequently, the method to be developed must adapt itself automatically to this constraint.

A first method for extracting letters from drop caps was described in [23]. This method is based on the study of shape descriptors. Instead of trying to find the best discriminating descriptor, the authors suggest to use a combination of several descriptors depending on the type of symbol to be extracted. They automatically assign to each descriptor a recognition map, based not only on the recognition rate but also on the distributions of the errors, to define a descriptor measure for each cluster of symbols. They applied the method to drop caps and to letter extraction but no quantitative evaluation was provided to quantify the quality of their extraction.

As presented in the previous section, lettrines are composed of different layers of information. Each layer can be defined by its specific spatio-frequential features. It will be relevant to use a method that takes into account the decomposition in layers.

Wavelets decomposition is a one-step process that can be used for splitting low frequencies from high-frequencies of image, at different scales [16]. In this context, between each scale, a smoothing function is applied, while a low-pass filter and a high-pass filter extract different frequencies. Since the letter corresponds to large homogeneous area with low graylevel variation, we can recover it from low-frequency layer. However, this method could not be used to extract letter. Indeed, boundaries of letters do not appear in low frequencies and the smoothing function erases thinner boundaries (see figure 4 to see an example of degraded image).

This discussion leads us to introduce our own strategy, based on Meyer decomposition, which enables to filter out the noise and to use spatial frequencies of lettrine images to segment them into separate layers: a shape layer, a texture layer, and a noise layer. Then, based on the resulting layer decomposition, the use of a Zipf law on the gray levels of the shape layer will allow us to detect large homogeneous areas, which correspond to the letter. Extraction of the letter then becomes feasible.



Original image | Meyer simplification | Wavelet simplification

**Fig. 4** Illustration of two kinds of image simplification effects on boundaries. From left to right: original image ; image obtained using Meyer decomposition ; image obtained unsing Haar wavelet

## 3 The grayscale Meyer decomposition

Decomposing an image into meaningful components appears to be one of the major goals of recent developments in image processing. This kind of model arises from a well known but ill-posed problem in image processing: the aim is to recover an ideal image $u$, from a degraded observation $f$, as follows: $f = Au + v$ (where $A$ is a linear operator representing blur and $v$ is the noise, often additive). A classical approach consists of introducing a regularization term leading the system to admit a unique solution. This minimization problem leads to a number of applications already successfully implemented, such as image restoration, deconvolution, deblurring, zooming, inpainting, classification, colorization, segmentation, and optical flow regularization.

The first goal was image restoration and denoising, but following the ideas of Yves Meyer [18], in the total variation minimization framework of L. Rudin, S. Osher and E. Fatemi [15], the decomposition of images into geometrical and oscillatory (i.e texture) components appears to be a useful and very interesting approach for our image-analysis problem. We want to obtain the primary structure of images in order to properly segment the drop cap, independently of its textured portions, and to avoid acquisition problems such as noise.

The images of drop caps are very complex, very rich images in terms of information, and need to be simplified. These images are mainly made up of lines, which are unsuitable for the usual texture methods. We therefore use an approach developed by Dubois and Lugiez [12,13] which relies on a series of projection to separate the original image into several layers of information.

The first one, $u \in BV$ [3], containing the structure of the image, a second one, $v \in G$ [4], the texture, and the third one, $w \in E$ [5], the noise. Meyer decomposition is an iterative process which extracts details of image. Signal of image is projected in particular spaces ($J$, $J^*$ and $B^*$) to only keep interesting parts. A residual part ($\frac{1}{2\alpha}\|f-u-v-w\|_{L^2}$) is present in the minimization of the functional to get all the elements that are not caught by the other projections. At the opposite of wavelet transformation approaches, boundaries are kept into geometric part and letter can then be better extracted by using a connected component approach. For better comprehension of differents spaces, see [3–5].

This decomposition model is based on a minimization of this discretized functional $F$:

$$\underset{(u,v,w)\in X^3}{inf} F(u,v,w) \tag{1}$$

with

$$F(u,v,w) = J(u) + J^*\left(\frac{v}{\mu}\right) + B^*\left(\frac{w}{\lambda}\right) + \frac{1}{2\alpha}\|f-u-v-w\|_{L^2}$$

where $J$ is the total variation related to the extraction of the geometrical component, $B$ is a norm defined on the Besov space, $J^*\left(\frac{v}{\mu}\right)$, $B^*\left(\frac{w}{\lambda}\right)$ are the respective Legendre-Fenchel transforms[6] of $J$ and $B$ [5] for the extraction of texture and noise components. Parameter $\alpha$ controls the $L^2 - norm$ of the residual $f-u-v-w$ and X is the discrete Euclidean space $\mathbb{R}^{N\times N}$ for images of size $N \times N$.

To minimize this functional, Chambolle's projection algorithm is used [4]. The Chambolle projection $P$ on space $\lambda B_G$ [7] of $f$ is denoted $P_{\lambda B_G}(f)$ and is computed by an iterative algorithm. This algorithm starts with $P^0 = 0$ and for each pixel $(i,j)$ and at each step $n+1$ we have:

$$P_{i,j}^{n+1} = \frac{P_{i,j}^n + \tau\left(\triangle div\left(P^n\right) - \frac{f}{\lambda}\right)_{i,j}}{1 + \tau\left|\triangle div\left(P^n\right) - \frac{f}{\lambda}\right|_{i,j}} \tag{2}$$

In [8] a sufficient condition ensuring the convergence of this algorithm is given: $\tau \leqslant \frac{1}{8}$. To solve (1), the authors propose the algorithm of figure 5. The residual part of the functional is thus included in the noise layer.

*Algorithm :*
At step n :

1. $u$ and $v$ have been previously computed, we estimate:

   $\tilde{w} = P_{\delta B_E}(f-u-v)$
2. then we compute:

   $\tilde{v} = P_{\mu B_G}(f-u-\tilde{w})$
3. and we finally obtain:

   $\tilde{u} = f-u-\tilde{v}-\tilde{w}-P_{\lambda B_G}(f-\tilde{v}-\tilde{w})$

This operation is repeated until :

$max(|\tilde{u}-u|, |\tilde{v}-v|, |\tilde{w}-w|) \leqslant \varepsilon$

**Fig. 5** Grayscale image decomposition algorithm

In [4], the authors replace $P_{\delta B_E}(f-u-v)$ by $f-u-v-W_{ST}(f-u-v,\delta)$ where $W_{ST}(f-u-v,\delta)$ stands for the wavelet soft-thresholding of $f-u-v$ with threshold $\delta$ defined by :

$$S_\delta\left(d_i^j\right) = \begin{cases} d_i^j - \delta sign\left(d_i^j\right) & if \quad |d_i^j| > \delta \\ 0 & if \quad |d_i^j| \leqslant \delta \end{cases} \tag{3}$$

where $d_i^j$ is the wavelet coefficient, $j$ the resolution and $i \in \{x, y, xy\}$.

For our experimentations, we used the default parameters of the Meyer decomposition [4,12]. Figure 6 shows the application of the grayscale decomposition model to an image.

### 3.1 Layers in details

The three layers extracted using the implementation in [12] of the Meyer decomposition (Aujol and Chambolle algorithm) can also be seen as follows:

- The Regularized Layer corresponds to the area of the image that has low-level fluctuations of gray level. This layer enables us to highlight geometry, which corresponds to shapes in the image. In the rest of this paper, we will call this layer the "*Shape Layer*".

[3] $BV(\Omega)$ is the subspace of functions $u \in L^1(\Omega)$ such that the following quantity, called the total variation of $u$, is finite:

$$J(u) = sup\left\{\int_\Omega u(x) div(\xi(x))dx\right\}$$

for any $\xi \in C_c^1(\Omega, \mathbb{R}^2)$, $||\xi||_{L^\infty(\Omega)} \leq 1$

[4] $G$ is the subspace introduced by Meyer for oscillating patterns. $G$ denotes the Banach space composed of the distributions $f$ which can be written $f = \partial_1 g_1 + \partial_2 g_2 = div(g)$ with $g_1$ and $g_2$ in $L^\infty(\Omega)$. With $\|f\|_{L^\infty} = \sup_{t\in[a,b]}|f(t)|$, on G, the following norm is considered:

$$\|v\|_G = inf\{\|g\|_{L^\infty(\Omega,R^2)}/v = div(g),$$
$$g = (g_1, g_2), |g(x)| = \sqrt{(|g_1|^2 + |g_2|^2)(x)}\}$$

[5] It enables to model oscillating patterns. Let $\dot{B}_{1,1}^1$ be the usual homogeneous Besov space then the dual space of $\dot{B}_{1,1}^1$ is the Banach space $E = \dot{B}_{-1,\infty}^\infty$

[6] The Legendre-Fenchel transform of $F$ is given by $F^*(v) = sup_u(\langle u, v\rangle_{L^2} - F(u))$, where $\langle .,. \rangle_{L^2}$ stands for the $L^2$ inner product [24]

[7] $\lambda B_G = \{f \in G/\|f\|_G \leqslant \lambda\}$

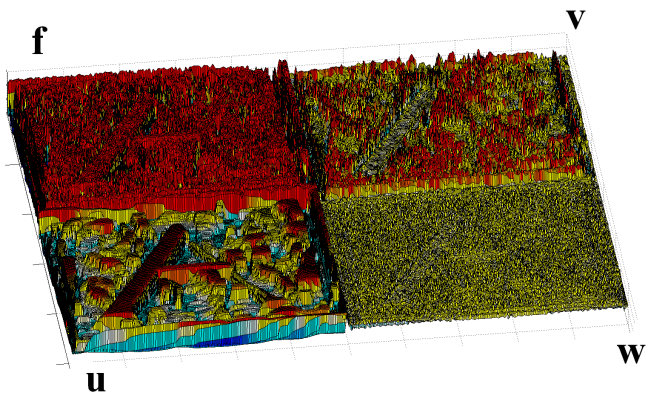



**Fig. 6** 3-D representation of image 7(a) decomposition. From top to bottom and left to right: original noisy image ($f$), regular or geometrical part ($u$), Texture component ($v$) and noise component ($w$).

- The Oscillating Layer, which corresponds to the oscillating element of the image. In our case, this layer highlights the texture in drop caps and in the rest of this paper we will call this layer the "*Texture Layer*".
- The highly Oscillating Layer, which corresponds to noise in the image. In fact, this layer recovers everything that does not belong to the first two layers. This layer therefore incorporates noise, background text, and the problem of aging paper or document. Our goal is to recognize images in the old document images while being robust towards noise variations. For this reason we do not refer to this layer in the next sections of this paper.

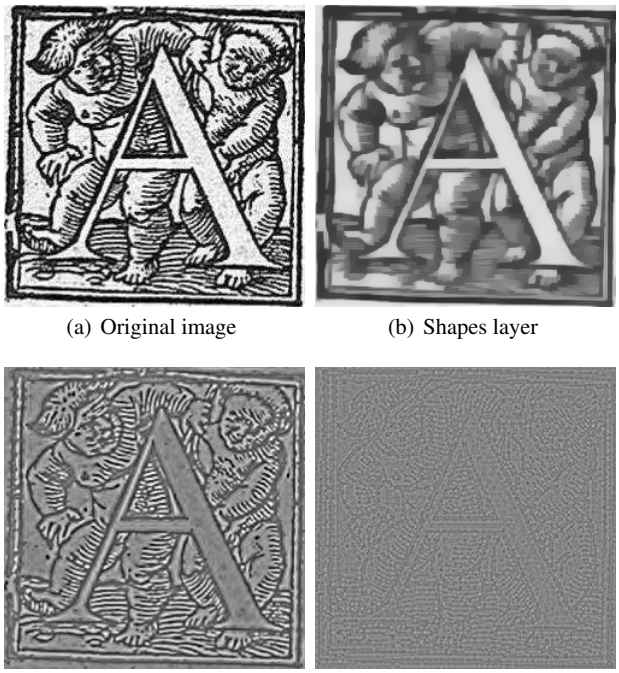
An example of decomposition is given in Fig. 7.
*Specific Treatment* Each layer will be seen as an image composed of uniform elements (the first layer consists only of shapes and the second consists only of textures). In the case of the regularized layer, we use a Zipf law to model the distribution of patterns.

## 4 Zipf's Law

### 4.1 Introduction

As mentioned, in the Meyer-based segmentation part, the shape layer issuing from this stage has to be considered for extraction of the letter. For this purpose, since component extraction is not adapted, we propose to use a complementary Zipf law based segmentation stage in order to extract large and homogeneous regions which correspond to the letter.

In order to become robust with respect to noisy conditions and detail changes, we chose a statistical approach based on the frequency of the patterns. The variety of colors involved in drop caps is considerable and never known. Besides, the letter is more than often in a single tone within an image. We then looked for a model to approximate the distribution of patterns present in the drop caps. This model is Zipf's law.



(a) Original image (b) Shapes layer

(c) Textures layer (d) Noise layer

**Fig. 7** An example of drop cap decomposition using Meyer decomposition

In this section we will review Zipf's law and its application to images, in particular to drop caps; in fact, a mixture of laws is observed. We then compare the results according to the nature of the image.

### 4.2 Zipf's Law

Zipf's law [26] is an empirical law formulated fifty years ago, which relies on a power law. The law states that in phenomena described 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 $M_1$, $M_2$...$M_n$, denoted $N_1$, $N_2$...$N_n$, are related to the rank of these symbols when sorted with respect to their frequency occurrence. The following relation holds:

$$N_\sigma(i) = k * i^a$$

$N_\sigma(i)$ represents the occurrence number of the pattern with rank $i$, and $k$ and $a$ are constants. This power law is characterized by the value of the exponent $a$; $k$ is more closely linked to the length of the symbol sequence studied. The relationship is not linear but a simple transform leads to a linear relationship between the logarithm of $N$ and the logarithm of the rank. The value of exponent $a$ can then easily

be estimated from the leading coefficient of the regression line approximating the experimental points of the 2D graph ($\log(i)$, $\log(N_\sigma(i))$) with $i$=1 to n. Below, the graph is called the Zipf graph. An example is given in figure 8. As one can see in figure 8, Zipf law graph can be approximated by a set of three linear functions. This point has been observed for all our experiments as we will see in the next part. If one wants to approximate these linear functions, one way to achieve the approximation is to use the least-squares method. Since the points are not regularly spaced, the points of the graph are re-scaled along the horizontal axis before approximation.

### 4.3 Image Decomposition - Layer Extraction

In this section, we note some problems that may arise with images. In the case of unidimensional data, the only possible tuples are limited to successive symbols. When images are concerned, they must be replaced by masks respecting the topology of the 2-D space in which the data are embedded. We have chosen to use 3x3 masks as the most common neighborhood of a pixel in a 2-D space.

The principle then remains the same: the number of occurrences of each pattern is computed. Nevertheless, since 256 symbols are used to code pixels, there could theoretically be $256^9$ different patterns. This number is much larger than the number of pixels in an image. Indeed, if all patterns are represented only once, no reliable model can be deduced, and the statistics will lose their significance. For example a 640x480 image contains only 304,964 patterns. It then becomes necessary to restrict the number of perceived patterns in order to give sense to the model. The coding is decisive in the matter.

The aim is to find the most suitable method of encoding to define indexes capable of distinguishing each component of a drop cap.

Some studies have shown that a Zipf law applies to the case of images with different encoding processes [7]. We are looking for a coding process that produces models capable of distinguishing the images we are studying. This qualifies as an effective coding process.

The drop caps we are studying have been scanned as gray-level images where each pixel is encoded with 8 bits (256 different levels). The intensity is the information encoded.

As indicated by our previous remarks, the number of different symbols must be reduced. Two ways to achieve the reduction are possible: either the number of symbols used to characterize the pixel is reduced or the number of the pixels involved in the mask is reduced.

Here our motivation is to preserve the pattern of a scene that relies more on the differences of gray levels than on their absolute values.

A simple approach would be to use only $k$ gray levels to characterize the intensity level of the pixels. Most often, a small number of gray levels is sufficient to observe an image keeping the significant details. As a quantization into $k$ equal classes would lead to unstable results, we have chosen to use a method for classifying the gray levels into $k$ classes by way of a k-means algorithm [17]. We experimented with various values of $k$ and decided to keep only three color levels. This is in accordance with the appearance of these historical images. In Fig. 8 we show an example of a Zipf curve obtained with the 3-means algorithm.

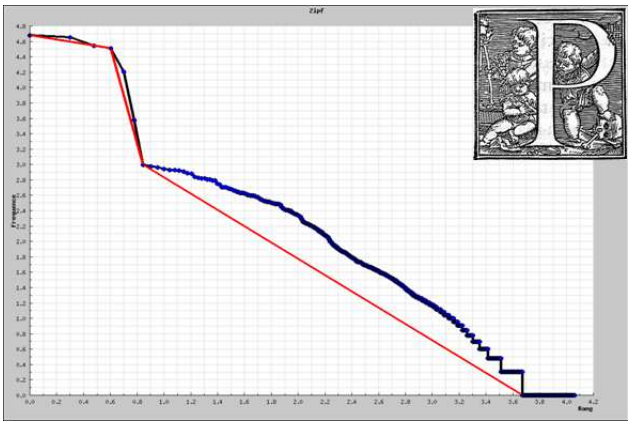


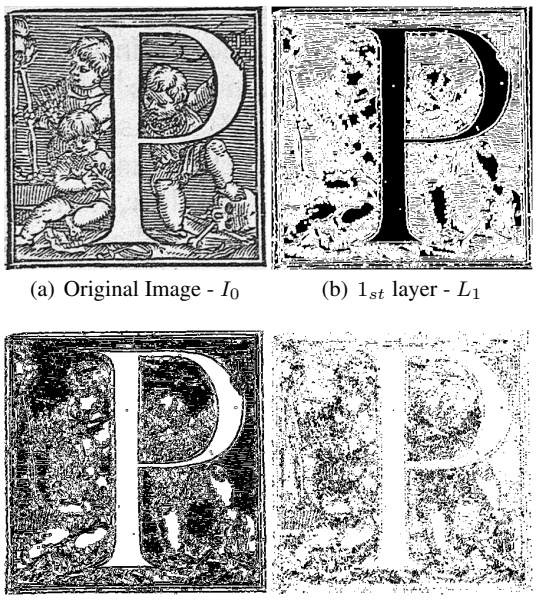
**Fig. 8** Example of a Drop Cap and its Zipf plot, showing the various straight zones extracted

A closer examination of the curves shows they are not always linear throughout, *i.e.* Zipf's law does not hold for the whole pattern set. Nevertheless, a number of straight-line segments can be observed. This means that several structures can be observed in the distribution of patterns. This is quite natural, since several things appear in a drop cap, at least the letter and the background, which may be very complex. A Zipf law cannot properly approximate the distribution, however a mixture of such laws may be considered. Depending on the coding process used, these zones (that is to say the corresponding zones in the image) may be interpreted. We observed that the left-hand portion of the graph was concerned with the regions in the image whereas the right-hand portion gave information on the contours present in the image. We can extract from both some structural information about the regions, and also the structure of the contours within the images. This can be called a Zipf decomposition of an image, as presented in (4), where $I_0$ corresponds to the original image and $L_n$ to the $n^{th}$ layer with $n \in [1;3]$. Here the sum is an exclusive sum. We obtain a partition of the image.

$$Z(I_0) = L_1 + L_2 + L_3 \tag{4}$$

Since letters are often very regular, uniform zones, the patterns they contain are the most frequent and the zone in the curve, associated with the letter is the left-most portion, however many zones are present. The various linear zones are automatically extracted as shown in Fig. 9, using a recursive process. The splitting point in a curve segment is defined as the farthest point from the straight line linking the two extreme points of the curve. Note that the image carries a mixture of several phenomena, which are highlighted in the expression. Several power laws are involved, from which several exponent values can be computed.

It is possible to recover from the image the pixels that are contributing to the various zones of the Zipf curve. We call them layers and the image is then decomposed into several layers. The patterns involved in the first layer (associated with the LH portion of the Zipf curve) are texture patterns whereas the patterns associated with the RH portion are contour patterns. An example is shown in figure 9. The image identified as the first layer consists of pixels obtained from the pattern of the first Zipf law, second layer with pixels from the second power law, and the third layer with pixels from the third Zipf law.



(a) Original Image - $I_0$ (b) $1_{st}$ layer - $L_1$

(c) $2_{nd}$ layer - $L_2$ (d) $3_{rd}$ layer - $L_3$

**Fig. 9** An example of three layers extracted from images. Each black pixel corresponds to a pixel in the original image that belongs to the layer.

The method can characterize not only the image's overall appearance but also its structural composition. We note that the first layer comprises the letter and large areas derived from the background. The second layer is made using thick outlines, and the third one, with thinner outlines. The first layer is the most interesting in terms of our present objective.

Our method is not driven by the gray level of the letter in the drop cap. In spite of the various possible colors of the letter, all the patterns belonging to the letter will be included in the first portion of the Zipf curve. This is one of the main reasons for deciding to use a model that relies on the frequency rather than on the overall aspect of the patterns.

## 5 Letter Extraction

In the two previous sections, we have discussed methods that can address the difficulties encountered in the extraction of letters from lettrines. Nevertheless, neither of the transforms provides an obvious criterion for extracting the letter.

Indeed we have two decomposition methods with two distinct objectives. One method considers the image as the superposition of three different signals (Meyer based approach), while the other considers the image as the juxtaposition of different objects that carry different meanings (Zipf law). Of course neither of these point of view is correct and we will benefit from the cooperation between the two methods. Applying the first decomposition allows to have a more pure signal before considering the partition of the image.

In case of the Meyer decomposition it is difficult to determine whether the letter is obtained from the lightest or the darkest pixels, but textures are separated from shapes. From the first layer modeled with the most frequent Zipf law one can separate the connected components of lettrines but they do not fit well with the letter, their area is often very small and the largest is not always the letter, owing to a lack of regularity in the original image. In what follows we attempt to coordinate the two approaches.

When the first layer of Zipf decomposition is extracted from the regularized layer of the Meyer transformation, the connected components are more regular and have a larger area. An example of Zipf decomposition result obtained from the original image and from the regularized layer of the Meyer transformation can be observed in figure 10.

The largest connected component extracted begin to acquire some semantic meaning, for example if the largest one covers a significant portion of a face or a letter. Nevertheless, the letter is not always the largest connected component of the first Zipf layer of the regularized layer in a Meyer transformation. In this case criteria have to be established for distinguishing between the letter and elements of the background. For reasons of legibility, the letter is a large element, it is centered in the drop cap and it does not touch the border of the drop cap. We are searching among the largest centered connected components that do not belong to the boundaries, in order to extract the letter. Fig. 11 presents some results progressing from the original image to the letter, displaying

(a) Original Image (b) Zipf law on original image (c) Zipf law on the regularized layer of the Meyer decomposition

**Fig. 10** An example of Zipf decomposition result obtained from the original image and from the regularized layer of the Meyer transformation. We can observe that the connected components extracted from the regularized layer of the Meyer transformation are more regular and they have a larger area. This allows to get a better visual understanding of image.

the image at each step (Meyer shapes layer and first Zipf layer obtained from the previous example).

## 6 Experiments and validation

Based on the appropriate sequencing of Meyer decomposition and Zipf law analysis, we developed a system allowing to extract the letter from the lettrines.

The evaluation of such a system is a fundamental step because it guarantees its usability to the users, and because it provides an objective view of the system. In the context of such a project, the implementation of an objective evaluation system is rather difficult, because of the variability of the users' requirements: research historians, and netsurfers, are likely to search for very different kinds of information.

In the context of the NaviDoMass project, and more specifically for the purpose of dropcap indexing, we decided to evaluate the quality of our system by considering the objective of "Letter Based Retrieval". This choice is motivated by the fact that many historians want to be able to retrieve drop caps in terms of this criterion. Accordingly, the evaluation of our system relies on the application of an OCR system at the conclusion of the letter segmentation. As a result, the classification rate is the main performance evaluation criterion for our system.

We conducted different series of tests in order to validate our choices. The first series of tests consist in highlighting the difficulties of letter extraction by comparing our results with a classical Otsu binarization process [19].

Both approaches are followed by an extraction and selection of connected components. Then we compare results obtained by two OCR on a set of lettrines. For the evaluation, we used a commercial OCR system (FineReader), as well as an open-source system (Tesseract). We applied our approach to an image database containing 4,293 images: 1,293 of these images were used for the training set, and 3,000 for the tests. The results are summarized in Table 1.

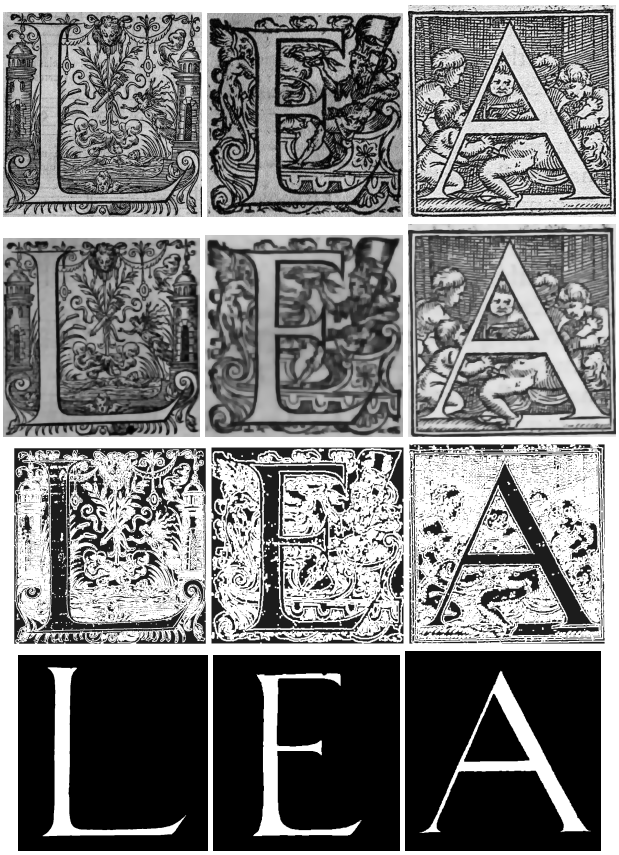


**Fig. 11** Examples of letters automatically extracted from drop caps. The first row corresponds to the original images, the second one to the shape layer of Meyer decomposition, the third row to the most frequent patterns extracted from the original images, and the last row to the connected component selected.

| Using Otsu criterion | | |
|---|---|---|
| | FineReader | Tesseract |
| Classification Rate | 20.76% | 22% |

| Using original images with Zipf law | | |
|---|---|---|
| | FineReader | Tesseract |
| Classification Rate | 48.2% | 39.6% |

**Table 1** Recognition rates for letters in lettrines using two different approaches of letter segmentation and two kinds of OCR

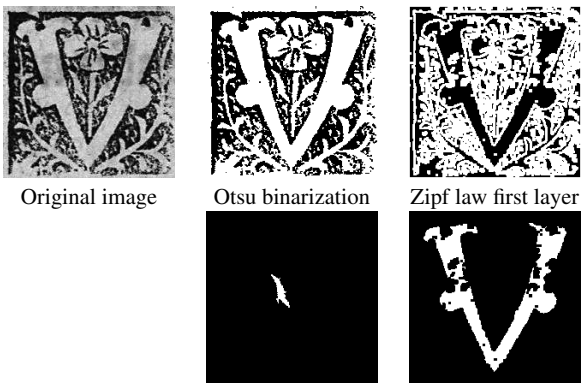
We observed that the recognition rates obtained by using the Zipf algorithms are two times better than those obtained using Otsu binarization. Figures 12 and 13 show some examples of letters extracted in order to illustrate our approach. We see in figure 12 that extraction of connected component requires to know a priori the color of the connected compo-

nent to be extracted. In the case of Otsu's approach, black letter can not be extracted . Thanks to Zipf law abilities for extracting most frequent patterns, our approach permits to dynamically adapt the letter extraction. As a letter is a wide uniform area, it is highlighted regardless of its color.

Original image Result using Otsu Result using Zipf law

**Fig. 12** Comparison between Otsu binarization and Zipf law segmentation. The main difference relies on the fact that Otsu method binarize image and black letters can not be extracted.

The second major difference: extraction of greater part of a connected component, is presented in Figure 13. In this figure, the same image has been segmented using Otsu binarization and Zipf law segmentation, and the letter extracted is presented. We can see in the original image that background is degraded and letter is connected to the outline of lettrine. With the Otsu binarization, letter stays connected to the outline, while Zipf law segmentation leaves out pixels of the link.



Original image Otsu binarization Zipf law first layer

Letter extracted Letter extracted

**Fig. 13** Comparison between Otsu binarization and Zipf law segmentation. Otsu binarization only associate each pixel with the background or foreground from its value. Zipf law segmentation relies on pattern frequencies, so large area like letter can be extracted even if it is connected to background (link correspond to pattern with low frequency).

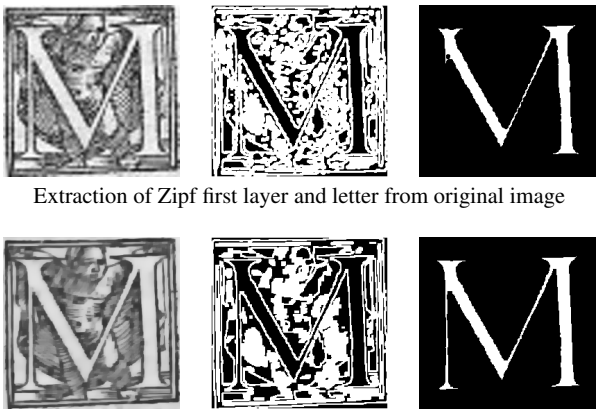
From this first series of experimentations, we deduce that Zipf law approach give better results for connected components extraction. We thus decided to use it in the second series of experimentation. These tests consisted in comparing Meyer decomposition to a wavelet approach to assess the better adequacy of Meyer decomposition to an extraction of connected component. Table 2 presents results obtained with the database and OCR of the first experiment.

| Using Haar wavelet with Zipf law | | |
|---|---|---|
| | FineReader | Tesseract |
| Classification Rate | 29.1% | 27.8% |

| Using Meyer's shapes layer with zipf law | | |
|---|---|---|
| | FineReader | Tesseract |
| Classification Rate | 72.8% | 67.9% |

**Table 2** Recognition rates for letters in lettrines using two different approaches of image simplification and two kinds of OCR.

As we predicted, the recognition rates obtained by using the combination of Meyer and Zipf algorithms are better than those obtained using Haar wavelet and Zipf algorithm. Figure 14 shows the same letter extracted using both approaches. At each level of wavelet decomposition, the smoothing function slims down boundaries and open regions. In the example presented in first line, letter M is split in many parts and cannot be properly extracted. In the case of Meyer decomposition, regularized image is obtained after a projection in a space that preserve boundaries while smoothing uniform areas. Letter extracted is then complete and recognized by OCR. Moreover, with Meyer decomposition, the original image is denoised and the variation of gray levels is reduced. The letter is more perfectly extracted and OCR is able to recognize it.



Extraction of Zipf first layer and letter from original image

Extraction of Zipf first layer and letter from Meyer shapes layer

**Fig. 14** Examples of letter extractions. First column corresponds to the image regularized, the second column correspond to image after applying the Zipf algorithm, and the third column correspond to letter extracted. Each row presents the result for each approach tested and the difference between two results explain differences in recognition rates

Finally, the cases for which our system fails correspond to very difficult images, where the letter is composed not of one but of many connected components. Separation into

several connected components can be explained by two phenomena: the degradation of the paper over time and the stylistic effects. Some examples of poor extraction results are shown in Fig. 15. The first case highlights a stylistic effect used to represent the letter, and the two other cases show degraded lettrines. To further develop this approach, we are considering groups of connected components that resemble the selected one, that are in its neighborhood, and that share similar features (for example the same principal orientation or the same width).

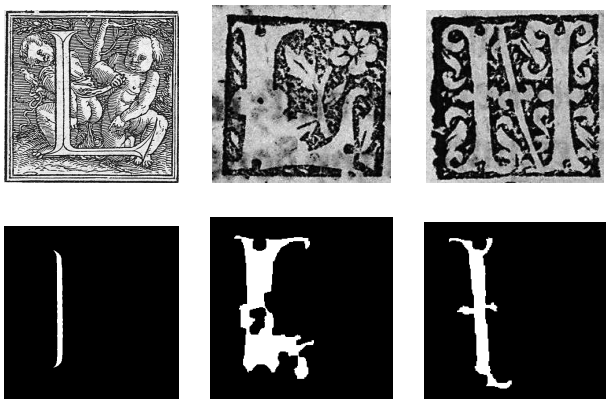


**Fig. 15** An example of very difficult letter extraction. The first line corresponds to the original image and the last line to the letter extracted.

## 7 Conclusion

In this paper we consider the problem of extracting letters from drop-cap images. Our strategy consists of decomposing the information into several layers, as a function of the identified features for each of these layers. Based on the frequency properties of the information, the Meyer-based decomposition allows a separation of the various layers, in which the information can be more easily analyzed. Then, based on the use of a Zipf modelling which enables the extraction of areas potentially corresponding to the letter, a letter extraction can be performed.

Evaluation of the overall process highlights the relevance of the approach, by enabling the extraction of letters from a significant set of data. Further development of this work may require the use of other layers (textures and noise), in order to complete the indexing process. In addition we will consider the use of visual keywords for the various layers, in order to achieve a comprehensive and practical indexing process.

## References


1. http://www.delos.info/.
2. http://www.dw-world.de/dw/article/0,1564,1566717,00.html.
3. J. F. Aujol, G. Aubert, L. B. Feraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, Jan. 2005.
4. J.-F. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, 2005.
5. J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.
6. J. Bigun, S. K. Bhattacharjee, and S. Michel. Orientation radiograms for image retrieval: An alternative to segmentation. In *International Conference on Pattern Recognition*, volume 7276, 1996.
7. Y. Caron, P. Makris, and N. Vincent. Use of power law models in detecting region of interest. *Pattern Recogn.*, 40(9):2521–2529, 2007.
8. A. Chambolle. Total variation minimization and a class of binary MRF models. *EMMCVPR*, 3757 of Lecture Notes in Computer Sciences:136–152, 2005.
9. H. Chouaib, F. Cloppet, and N. Vincent. Graphical drop caps indexing. In *GREC*, pages 212–219, 2009.
10. M. Coustaty, J.-M. Ogier, R. Pareti, and N. Vincent. Drop caps decomposition for indexing a new letter extraction method. In *10th International Conference on Document Analysis and Recognition*, pages 476–480, Barcelona, Spain, 2009. IEEE Computer Society.
11. M. Delalandre. Retrieval of the ornaments from the Hand-Press period: an overview. In *Internation Conference on Document Analysis and Recognition*, volume 2, pages 496–500, Barcelona, Spain, 2009.
12. S. Dubois, M. Lugiez, R. Péteri, and M. Ménard. Adding a noise component to a color decomposition model for improving color texture extraction. In *4th European Conference on Colour in Graphics, Imaging, and Vision*, pages 394–398, 2008.
13. A. E. Hamidi, M. Menard, M. Lugiez, and C. Ghannam. Weighted and extended total variation for image restoration and decomposition. *Pattern Recognition*, 43(4):1564 – 1576, 2010.
14. N. Journet, J.-Y. Ramel, R. Mullot, and V. Eglin. Document image characterization using a multiresolution analysis of the texture: application to old documents. *IJDAR*, 11(1):9–18, 2008.
15. L.Rudin, S.Osher, and E.Fatemi. Nonlinear total variation based noise removal. *Physica D*, 60:259–269, 1992.
16. S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3 edition, Dec. 2008.
17. J. B. McQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967.
18. Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*. The fifteenth dean jacqueline B. Lewis Memorial Lectures, 2001.
19. N. OTSU. A threshold selection method from Gray-Level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.
20. R. Pareti, S. Uttama, J. Salmon, J. Ogier, S. Tabbone, L. Wendling, S. Adam, and N. Vincent. On defining signatures for the retrieval and the classification of graphical drop caps. In *Second International Conference on Document Image Analysis for Libraries*, pages 220–231. IEEE Computer Society, 2006.
21. R. Pareti and N. Vincent. Ancient initial letters indexing. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 756–759, Washington, DC, USA, 2006. IEEE Computer Society.
22. C. Remazeilles. *Etude des processus de degradation des manuscrits anciens ecrit a l'encre ferrogallique*. PhD thesis, La Rochelle, 2001.

23. J. P. Salmon, L. Wendling, and S. Tabbone. Improving the recognition by integrating the combination of descriptors. *Int. J. Doc. Anal. Recognit.*, 9(1):3–12, 2007.
24. J. L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. Image Processing*, 14(10):1570–1582, Oct. 2005.
25. S. Uttama, P. Loonis, M. Delalandre, and J. Ogier. Segmentation and retrieval of ancient graphic documents. In *Graphics Recognition. Ten Years Review and Future Perspectives*, pages 88–98. LNCS - Springer Verlag, 2006.
26. G. Zipf. *Human Behavior and the Principle of Least Effort*. Hafner Pub. Co, 1949.