



A complete pyramidal geometrical scheme for text based image description and retrieval

Guillaume Joutel, Véronique Eglin, Hubert Emptoz

► To cite this version:

Guillaume Joutel, Véronique Eglin, Hubert Emptoz. A complete pyramidal geometrical scheme for text based image description and retrieval. International Conference on Image and Signal Processing 2008, Jul 2008, Cherbourg-Octeville, France. pp.471-480, 10.1007/978-3-540-69905-7_54 . hal-01501888

HAL Id: hal-01501888

<https://hal.science/hal-01501888>

Submitted on 4 Apr 2017

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.

A complete pyramidal geometrical scheme for text based image description and retrieval

Guillaume JOUTEL, Véronique EGLIN, Hubert EMPTOZ
LIRIS UMR CNRS 5205 – INSA Lyon, 69621 VILLEURBANNE Cedex
E-mail: {guillaume.joutel; veronique.eclin@insa-lyon.fr}

Abstract. This paper presents a general architecture for ancient handwriting documents content description and retrieval. It is based on the Curvelets decomposition of images for indexing linear singularities of handwritten shapes. As it belongs to the Wavelets family, its representation is used at several scales of details. The proposed scheme for handwritten shape characterization targets to detect oriented and curved fragments at different scales: it is used in a first step to extract visual textual interest regions and secondly to compose a cross-scale signature for each handwritten analyzed samples. The images description is studied through different kinds of deformations that show the efficiency of the proposition for even degraded and variable handwriting text. The complete implementation scheme is validated with a content based images retrieval (CBIR) application on the medieval database from the IRHT¹ and on the European 18th century correspondences corpus from the CERPHI.

1 Introduction

1.1 Digital Paleography and image analysis

In this work, we are interested in digitized Middle-Age (composed by copyists' texts from the 9th to the 15th century) and Humanistic manuscripts (essentially composed by authors' drafts from the 18th and 19th century) analysis. It is dedicated to palaeographers and historians to help them in their everyday work of manuscripts dating, expertise and authentication, see figure 1. One of the primary difficulties faced by palaeographers is the classification and identification of hands, and this is an area which has already received a good deal of attention in other disciplines. Specifically, the community of forensic document analysts have been working for several years now to develop computer-based systems for retrieval, identifying and classifying modern handwriting, and this begs the question whether such work can be applied to medieval writing as well. In that context, we need an original approach to answer the opposite questions raised by those two kinds of documents from the *medieval* and the *humanistic* collections. In the first case, the objective is to find writer independent and style dependent primitives for Middle-Ages manuscripts page images retrieval, while for humanistic manuscripts, the objective is to find writer dependent primitives for an identification task. The answers to those objectives are complex. We propose in this paper a contribution to this challenge that must be considered as a tool kit for historians and medievalist palaeographers to help them in their work of manuscripts dating, expertise and authentication. Due to the fact that palaeographers judge that the *curvature* and the *orientation* are two fundamental dimensions of handwritings, we have searched a way to compute them on the shapes for both collections. To compute those dimensions, we have developed a methodology that is sensitive to variations at the frontiers of

¹ This study has been supported by two projects: the ANR GRAPHEM 2007-2010 project between the IRHT (Institut de Recherche en Histoire des Textes), the Ecole des Chartes from Paris and the LIRIS and the regional Cluster 13 Project: "Inheritance, culture, Creation", 2005-2009 in collaboration with the CERPHI of Lyon (Centre d'Etude en Rhétorique, Philosophique et Histoire des Idées).

shapes, that pays attention to the evolution of these criteria at different scales, that reveals the variability of shapes and their anisotropy, that is robust to disturbed environments (presence of disturbed backgrounds, of partial shapes...) and finally, that doesn't require prohibitive costs of treatment nor great storage volume for each analyzed page image. Our choice relates to a redundant multi-scale transform: the Curvelets transform that is more robust than wavelets for the representation of shapes anisotropy, of *lines segments* and *curves* in the images. The proposed solution lies for a great part on an original indexing and navigation system that contributes to different objectives: handwritten documents retrieval from the same writer, writers' identification, middle-age handwriting clustering.

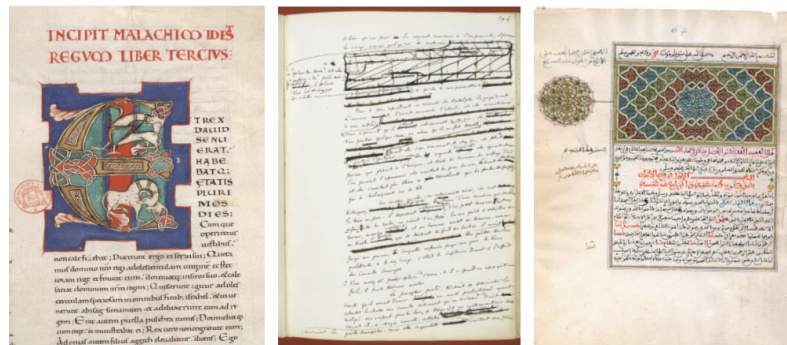


Fig. 1. Complex handwriting documents images from different periods (13th and 18th century)

I.2 A methodology based on coarse interest regions selection and fine shape description

Our CBIR system is based on geometrical multiscale images decomposition (the Curvelets) embedded in the different stages of the architecture. The figure 2 presents the system architecture into two main parts.

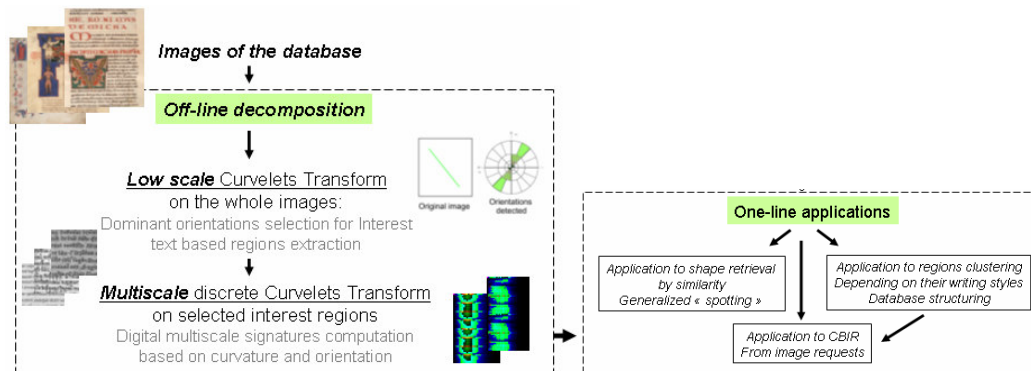


Fig. 2. Pyramid architecture for the handwriting characterization and on-line applications.

The first part presents an off-line Curvelet based decomposition that lies on the exploitation of the distributions of Curvelet strong coefficients through different scales: the lowest scales for the estimation of interest text areas, the middle and highest scales for the computation of digital signatures of homogeneous text fragments. The digital signatures are the 2D representations of all couples of values (curvature, orientation) obtained by the Curvelets decomposition. Three applications have then been investigated by exploiting this off-line decomposition:

- The *writing styles clustering* that is based on the production of a cross-scale features vector for each handwritten samples and that contains a set of statistical measures from the different multiscale coefficients distributions. This clustering is based on an unsupervised classification scheme that groups palaeographical handwritten styles into visual coherent handwriting families.
- The *spotting* and the *shape retrieval by similarity* that lies on the analysis of the curvature and the orientation from the finest scale of the decomposition of the image.
- The *CBIR* that is also based on the finest scale analysis and on the production of an unique and characteristic digital signature for each handwritten samples.

In this paper, we focus on the application of CBIR that is based on a direct comparison between a request image, considered by its digital signature and all off-line signatures for the rest of the database images. The similarity criterion is a weighted value of correlation between the request signature and each individual signature of the images of the database.

1.3 Content based information retrieval for handwriting images databases

The simplest scenario of content based image retrieval is the one of global example-based search: the user chooses an image example and the system determines the images of the base with the most similar visual appearance. The principle of this approach has been established by Ballard in 1991, [Ballard]. It is the fundamental principle of a lot of systems that deal with natural images, like QBic from IBM (Flickner 1995), PhotoBook from the MIT (Pentland 1994), MARS (Multimedia Analysis and Retrieval System), PicToSeek, Ikona and the KIWI system, [10]. All those systems are sharing the same kinds of features for a global content based document images retrieval (colors, shapes and textures). In the specific field of CBIR dedicated to textual (printed or handwritten) document images, different specific (structural, geometrical or statistical) features and similarity measures have also been proposed, [9], [11]. But all of them have a common point: they are focused on the same property of images: the information is concentrated in linear shapes and their contours.

II Different ways to characterize handwritten shapes

II.1. Related works

The need for searching scanned handwritten documents are involved in application such as collections dating, works genesis, erudite study, critical edition, documents authentication... Recently Srihari et al. in [15-16] have realized the importance of handwritten document retrieval and have presented their retrieval system that is dedicated to forensics applications such as writer identification. Different distances are currently used to access the best matching between different set of handwriting samples, [7]. Discriminability of a pair of writing samples based on similarity value can be observed by studying their distributions when the pair arises from either the same writer or from different writers. Generally, in most writers' classification approaches, authors try to produce a set of exhaustive graphical characteristics which can efficiently describe all handwritings families. In handwriting classification which is our main goal, the most closely related works are [8] and [16]. In [8], authors propose an analysis of the variability of handwritings on the bases of two kinds of observations: firstly the thickness of the tracing and the spatial density of characters and secondly the successive directions of the tracing. This work has been led on contemporary documents only. The only work relative to the analysis of the Middle-ages documents referenced in [1] proposes classifications procedures which are today debatable by palaeographers. Some recent studies have tried to develop more robust approach for a non supervised classification. In [11] for example, authors are interested

in ancient Latin and Arabic manuscripts of the Middle-Ages before the emerging of the printing. They made the choice to analyze statistically the whole image of a manuscript and measure globally all patterns. This approach should guarantee the independency from the text content, the writer's personal style; the language used and the letters frequencies. It is difficult to pretend to be exhaustive in the description of handwritten shapes for the retrieval, so it is essential to work with expert users who are able to validate the measures that appear to be the most relevant. The signification degree that is assigned by the user could also guide the system to create a suitable distance between writing classes. In any cases, each computed measure should be evaluated in relation with all others. In this case, a Principal Component Analysis (PCA) considerably reduces the dimensionality of micro-features vectors, [13]. Within this kind of generic approach, it is possible to classify handwriting samples into visual distinct style classes. General methods are either based on the consideration of local particular graphemes [3] or on a too macroscopic and general characterization that is not always efficient for a complete writers' authentication, [5], [11]. In our proposition, we lead a Curvelets based analysis for a designed decomposition in curvature and orientation which have a real sense for palaeographers. Our contribution differs from the cited works because it is dedicated to both Middle-Ages and contemporary handwritings documents. Moreover it uses a geometrical multiscale decomposition: the Curvelets, which has never been used in this context.

II.2. A new interesting way: the geometrical decompositions

In the handwriting images, the discontinuities points are essentially gathered on the shapes' outlines that contain the maximum of the visible information. It is an information that can be analyzed both globally (orientations of outlines, presences of particularly striking and redundant curvatures), and locally (the discontinuities are situated in very sensitive locations in shapes contours where experts pay generally their attention to recognize or identify a writing). A mixed approach that combines global and local shapes description seemed to be an interesting way for handwriting images analysis. But we have noticed that all decompositions can not decorrelate both local and global images properties with the ability to localize all discontinuities in the image space. Fourier Transform is a powerful tool for analysing the global regularity of a function but it does not allow recognizing discontinuities points. It doesn't allow a local shape analysis and doesn't take advantage of multiscale correlations that exist in lines based images. Orthogonal wavelets provide frames that allow sparse representation (with only few coefficients) of images but present also a lack of directional characteristics for regular curves description and geometry. We are precisely interested here in the families of redundant transforms which allow to decompose the image into directed sub-bands and to capture the geometry in the shapes contours by the agglomerates of coefficients. We can give as example, the X-lets (*i.e.* contourlets, bandelets, complex wavelets and oriented wavelets. Among this family, we have chosen one specific X-lets: the Curvelets that have been introduced by Candès. They offer a set of the properties of directionality and good joint location of the shapes geometry. Ridgelets and Curvelets transforms that we defend here have been imagined to bypass wavelets disadvantages.

II.3 The Curvelet Transform - from continuous to discrete

General Properties of the continuous Curvelet Transform. The Curvelet transform is a multiscale multi-orientation transform with indexed atoms in location, scale and direction. The figure 3 shows the anisotropic Curvelet transform of Candès [6] of an image of our palaeographical database. On this figure, the transformation contains 4 scales with all coefficients, each one being of the same size as the original image. The combination of all images with all

missing scales (not represented here) reproduces with redundancy the original image. A study of the Curvelets coefficients properties, shows that the high valued coefficients tend to gather around the objects corners and contours in the image, what seems to be natural because there is a strong dependence between Curvelets coefficients from a scale to the other and for a given neighbourhood for various given scales and orientations. These dependences have been quantitatively verified by measuring the mutual information, [4].



Fig.3: Curvelet Transform of an image of the Palaeographical database (IRHT-Paris).

Conceptually, the Curvelet transform is a multi scale nonstandard pyramid because Curvelets have geometric features that set them apart from wavelets and the likes. Curvelets obey a parabolic scaling relation which says that at scale 2^j , each element has an envelope which is aligned along a ridge of length $2^{j/2}$ and width 2^j . Mathematically, if one works in \mathbb{R}^2 , we have first to consider a radial window $W(r)$ and then an angular window $V(t)$ where r and t are polar coordinates in the frequency domain. These are both smooth, nonnegative and real-valued, with W taking positive real arguments and supported on $r \in [1/2, 2]$ and V taking real arguments and supported on $t \in [-1, 1]$. These windows will always obey the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (3/4, 3/2); \quad \sum_{\ell=-\infty}^{\infty} V^2(t - \ell) = 1, \quad t \in (-1/2, 1/2)$$

Then, for each $j \geq j_0$, a frequency window U_j is defined in the Fourier domain by

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor} \theta}{2\pi}\right).$$

where $\lfloor j/2 \rfloor$ is the integer part of $j/2$. The support of U_j is a polar wedge with the support of W and V , the radial and angular windows applied with scale-dependent window widths in each direction. In Fourier space, Curvelets are supported near a “parabolic” wedge, and the shaded area represents such a generic wedge, see figure 4.

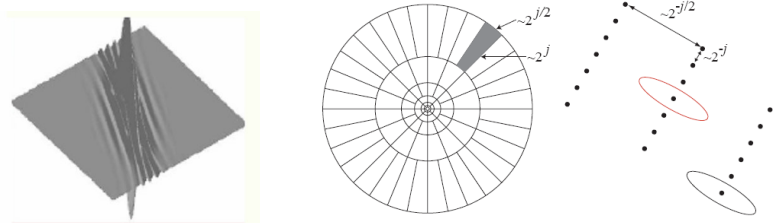


Fig.4 Elongated anisotropic support of the Curvelets defined by its parabolic dilation in polar frequency domain coordinates. Curvelet tiling of space and frequency. Representation of the increasing number of angular sectors with scale changing.

Procedural definition and implementation of the discrete version. The proposed Curvelets decomposition uses a combination of reversible operations: a sub-bands decomposition of the original 2D signal followed by a regular and normalized partitioning of images for each sub-band and then the application of a local Ridgelets transform on each partition.

We now describe a strategy for realizing a digital implementation of the Curvelet transform. The proposed implementation of the Curvelets decomposition has been suggested by Candès and Starck and constitutes the most common version. The successive simplified steps of the algorithm are listed below:

- 1/ Do the FFT of the initial image I
- 2/ Resample the frequential Cartesian space into recto-polar coordinates
- 3/ For each radial indexed \cdot_n line do the inversed Fourier Transform: so as to obtain the Radon Transform $\{\text{Rad}(\cdot_n, t_n)\}_{n=1..N}$ for each \cdot_n angle and then realize the wavelet 1D transform in the spatial domain.
- 4/ The inverse Ridgelet transform realizes all successive inversed transforms on each step.
- 5/ Initialize the subband decomposition for the Curvelet transform step.
- 6/ Compute a local Ridgelet transform on each subband.

In practice, we have privileged the new wrapping-based curvelet transform proposed by (Candès, 2002) that simplifies the calculation and increases the redundancy of the coefficients

III Off-line decomposition for the CBIR application

Interest textual regions selection. In textual regions of document images, lines orientations are globally regular near to the horizontal. This orientation can be detected at a low scale : the map of textual information is revealed by high Curvelets coefficients while other coefficients have been ignored. Of course, at this resolution of image, we have lost a part of precision in this detection that has been easily compensated by the detection of the *background map* that has been introduced in Ramel's work in [12].

Signatures construction, evolution and comparison. The Curvelet transform gives us an analysis of pixels for several scale and several orientations. We only focus here on the highest scale of the decomposition and only cumulate information relative to shapes contours. The principle of curvature estimation is described as follows: one pixel on a curve can be potentially detected in several orientations depending of the curve. Each Curvelets coefficient corresponding to this pixel in a detected orientation is then compared to all other detected orientations of the same pixel. We only retrieve the more significant coefficients values. Finally, the number of significant orientations gives us an evaluation of the overall curvature of a pixel, as it has been presented in [2] by J-P Antoine. From the orientations recovery and curvatures computation, we define a signature for each handwriting fragment. The signature is defined as the matrix of couple occurrences (curvature, orientation). Because we do not want to normalize our images before the decomposition, we have searched a way to normalize the signature. To do so, we keep in the normalized signature the ratio between every value in the original signature and the global amount of information of this signature. The only difficulty with this definition is that some coordinates in the signatures are common to almost all handwritings (presence of high values in the horizontal and vertical directions): we compensate this phenomenon by reversing ratios in the signature, i.e. by accentuating weakly represented couples (orientation, curvature). Figure 5 presents an example of matrix plotted as a pixel image for a Middle-Ages and a 18th author handwriting image. In our retrieval system, the distance between the queried signature and each signatures of the database is calculated using a normalized correlation similarity. The similarity measure $S(X,Y)$ between two image X and Y is defined as follows: $S(X,Y)=\text{Cov}(X,Y)/\sigma_x\sigma_y$, where $\text{cov}(X,Y)$ is the covariance between X and Y and σ_x and σ_y the standard deviations.

In order to know how our signature reacts to several possible deformations in a handwriting manuscript, we have led five kinds of tests: the horizontal and vertical stretching of text, the resolution changing, the letters' dimensions zooming and the lines' spaces changing. To have significant results, we have chosen to keep the same image dimensions after each deformation.

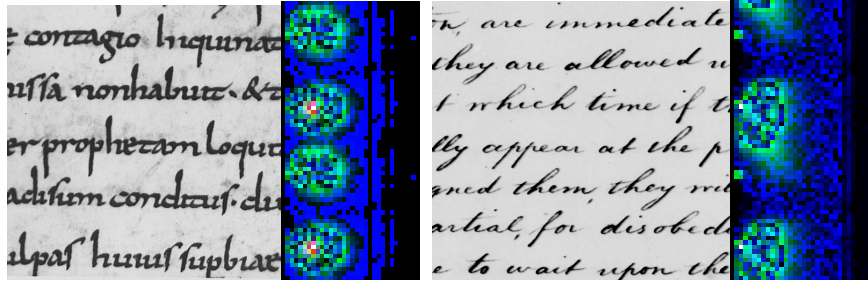


Fig.5. Signature of a Middle-Ages handwriting samples of the IRHT database and on a fragment extracted from the Georges Washington's digital collection.

The only general conclusions are presented here. Firstly, studying (vertical or horizontal) stretching, the signature's behaviour is exactly what we were expecting: i.e. a general shifting of values of couples (curvature, orientation) towards the orientation of the stretching (0° for horizontal and 90° for vertical), a general decreasing of curvatures and a great fall of the correlation coefficients between the original image and its successive deformations. In fact, the horizontal and the vertical directions are the two main dimensions in Latin handwritings. So modifying those dimensions led to great changes in the signatures design. Concerning the resolution and the letters' dimensions changes, we can take advantage on the presence of different scales in the decomposition of the Curvelets that stabilize the signature. The Curvelet transform offers the possibility to choose the best scale for the signature design depending on the initial image resolution and characters size. Lastly, modifications of the spaces between lines entail real consequences on the signature only when the amount of information is no more comparable. In an image with about ten lines of medieval text, if one removes or adds two or three lines, the rate of recognition is maintained.

IV The CBIR application: experimentation and quantitative evaluation

In this part, the test setup and the experimental results obtained for the image retrieval task are described. The images have been divided into two groups: one group is composed by known handwriting images and the second group is composed by queried images for testing. We have tried to test our retrieval system on two databases (over 800 images): the humanistic and the Middle-Ages databases (composed by manuscripts of the database of the IRHT and of European 18th century correspondences corpus (CERPHI)). In retrieval system, the performance is subjective and its evaluation is difficult, especially for the Middle-Ages base where we do not have precise information concerning Middle-Ages European handwriting dating process. The set of known images is selected and the image retrieval process is carried out against this set of known handwritings. In each case, we compute the average precision and the recall values for each database.

As we can see on figure 6.1 (middle age) the P/R curve tends to stabilize around a precision rate of 0,6 which is not a very high value but we have proceeded by unfavourable deductions that consists in counting as false answers all images that are not surely recognized as being similar to the request. This is due to the lack of certain ground truth about classification of the middle age database. The classification of our base is still in progress by palaeographers of the IRHT. On the other hand, on the figure 6.2 (humanistic), we can see that our system has exactly the same behaviour that we expected. In fact those rates are certainly the best that we can have, compared to other CBIR systems.

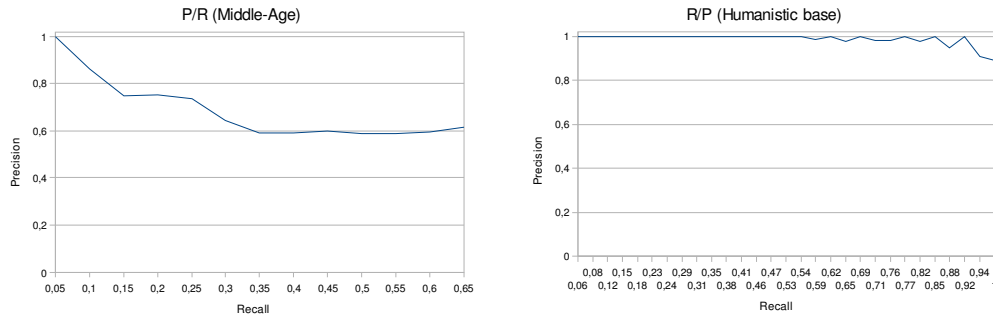


Fig.6. Average precision and recall curves in the middle-age (6.1) and the humanistic (6.2) database.

To have a better evaluation of our system, we have used a F-measure which combines the precision and the recall values ($F=2P.R/(P+R)$) and provides a single measure of the retrieval accuracy: it is the Harmony mean of precision and recall value. For the humanistic base, in the top 15 ranks, the F-measure is 96,61 ($P=1$ and $R=93$). This implies that considering the top 15 results supports high precision and recall measure. For the Middle-Ages base this value is 51,43.

Conclusion

We can say with confidence that Curvelets Transforms can be used as a general technique for feature detection. In the specific case of handwritings that we study, the hypothesis that curvature and orientation are fundamental dimensions of shapes seems to be validated by our CBIR system and its recall values. The system is currently integrated to the platform of the Graphem Project dedicated to paleographer experts. The two other on-line applications that are also based on the Curvelet geometric decomposition (spotting and writing styles clustering) are currently under development.

References

1. F. Aiolfi, M. Simi, D. Sona, A. Sperduti, A. Starita, G. Zaccagnini, "SPI: a System for Palaeographic Inspections. AIIA Notizie", p.34-38, Vol. 4, 1999.
2. P. Antoine, L. Jacques, "Measuring a curvature radius with directional wavelets", Inst Phys Conf Series, p. 899-904, 2003.
3. Bensefia, L. Heutte, T. Paquet, A. Nosary, "Identification du scripteur par représentation graphèmes", CIFED'02, p.285-294, 2002.
4. L. Boubchir, J. Fadili, "Bayesian Denoising Based on the MAP Estimator in Wavelet-domain Using Bessel K Form Prior", in Proc. of IEEE ICIP'2005 ; the IEEE International Conference on Image Processing, Vol I, pp. 113–116, September 11-14, Genoa, Italy, 2005.
5. M. Bulacu, L. Schomaker, "Writer style from oriented edge fragments", Computer Analysis of Images and Patterns (CAIP), p. 460-469, 2003.
6. E. Candès, D. Donoho, "Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges", in Schumaker L., *Curves and surfaces filtering*, Vanderbilt University Press, 1999.
7. S.H. Cha, S. Srihari, "Multiple Feature Integration for Writer Verification", IWFHR VII, p. 333-342, 2000.
8. J.-P. Cretz, "A set of handwriting families: style recognition ", ICDAR 95, p.489-494, 1995.
9. Journet N., Mullot R., Ramel JY, Eglin V. Ancient Printed Documents indexation: a new approach. . ICAPR'05. Third International Conference on Advances in Pattern Recognition, Pattern Recognition and Data Mining. Bath, United Kingdom.p. 513-522.Aout2005.
10. Loupias (E.), Bres (S.), Key Points based Indexing for Pre-attentive Similarities: The KIWI System , Pattern Analysis and Applications, Special Issue Image Indexing, summer 2001.

- 11 I. Moalla, F. LeBourgeois, H. Emptoz, A. M. Alimi, "*Contribution to the Discrimination of the Medieval Manuscript Texts*", Lecture Notes in Computer Science, Vol. 3872, p.25-37, 2006.
- 12 JY. Ramel, Leriche S., Demonet ML, Buisson S. User-driven Page Layout Analysis of historical printed Books. International Journal on Document Analysis and Recognition. Vol. 9. No. 2-4. p. 243-267. Avril 2007.
- 13] Shen, C., Ruan, X.G. and Mao, T.L., Writer identification using Gabor, 2002, Vol. 3, 2061-2064. [8, 13, 24]
- 14 Michael J. Swain, Dana H. Ballard, Color Indexing, International Journal of Computer Vision, 7:1, 11-32 (1991)
- 15 Zhang, B., Srihari, S.N., Binary vector dissimilarity measures for handwriting identification, in Document recognition and Retrieval, SPIE, 5010 pp.28-38, 2003.
- 16 Zhang, B., Srihari, S.N., Word image retrieval using binary features, in Document recognition and Retrieval, SPIE, 5296, pp.45-53, 2003.