We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Generic scale-space architecture for handwriting documents analysis

Joutel Guillaume, Eglin Véronique and Emptoz Hubert
Université de Lyon (LIRIS, UMR5205)
France

1. Introduction

Those last years, one can observe that the number of digital libraries has clearly increased but not as fast as the number of tools to promote the cultural heritage stored in those libraries. Among the great diversity of cultural inheritage we can notice that handwritten documents constitute an important part of ancient collections to valorise. In those document images, we are not interested in handwriting recognition and semantic content, but we focus on image content retrieval through visual low level characteristics of shapes (from graphemes to entire handwriting samples).

In this work, we are interested in digitized Middle-Age (composed by copyists' texts from the 9th to the 15th century), Humanistic manuscripts (essentially composed by authors' drafts from the 18th and 19th century) and overall lines based images, see Fig. 1.

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

The community of handwriting document analysis has been working for several years on the devlopment of computer-based systems for retrieval and identify handwritings from various historical periods. This begs the question whether such work can be applied to

medieval writing as well. In this work, on strong constraint is to find a robust characterisation tool for writer dependent or style dependent primitives

Due to the fact that the *curvature* and the *orientation* are judged to be two fundamental dimensions of handwritings, we have searched a way to compute them on the shapes for different corpus. To compute those dimensions, we have developed a methodology that is sensitive to variations at the edges of shapes and that reveals the variability of shapes and their anisotropy at different scales. The proposed methodology has been tuned so as to be robust to disturbed environments (presence of disturbed backgrounds, of partial shapes…) without requiring prohibitive costs of process 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. Standard wavelet-based tools suffer of their incapacity to locate thin structure of lines and thin variations all along curves. Those points will be discussed in following sections.

In the literature, we have observed that mainly handwritings characterization methods may be divided into four classes:

_ The first class gathers methods that search for structural information. As examples, we can quote the average height, the width and the overall slant of the writing. Despite these methods are really simple, their efficiency in the writer identification task is not proved as far. Their main drawback is in the initial segmentation task which is rarely robust.

_ The second class is made of statistical methods. The idea behind is to accumulate information in order to be less sensitive to intrinsic variations of writings. This time the drawback is the difficulty to evaluate quantitatively the amount of information that are needed in those kinds of process.

_ The third class is made of frequency analysis that is based on a change in the representation of the signal. This time the drawback is the lack of directionality that is caused by most of multiscale frquences based representation.

_ The fourth class tries to get the best of several methods by mixing classes. Currently, we may find in the fourth class some serial arrangements of two existing methods from the first three classes.

We think that, thanks to recent works on signal processing, it is possible to mix the three first classes of methods in a single and efficient characterization approach. This mixing process is proposed and discussed in parts 0 and 0. In those sections, we will show how the curvature and the orientation of shapes are essential features that are dependent on image frequences, geometry and ditribution. The choice of these features comes from several pluridisciplinary discussions with palaeographers that are working with. In the last part of the paper (in part 0), we will see two examples of applications based on these features: a content base image retrieval application and the analysis of the evolution process of Middle-Ages handwritings.

2. In the frequency domain

Most of the time, the structure of an image is very complex, you can find textured parts with changes more or less salient in brightness, or you can also encounter more or less overlapping objects. In all those situations it is not easy to define relevant universal features to describe objects in an image. In that context, objects are mostly defined according to

various criteria such as contour, shape, texture, presence of edges more or less smooth … It is clear that a unique representation fails to take into account all these characteristics and other models should be developed to optimise the image processing that need to be done. Among these transformations we can include coding, analysis, representation or segmentation of information. These new models of representation must take into account the most appropriate information of the image without looking for secondary ones.

If these new models can take into account the most appropriate information, it is still often convenient to be able to reverse the process. Therefore, we intend to define an invertible transform to work on information not limited to greyscales. The linear transforms of Fourier and Gabor were until recently the only alternatives to the representation and modelling of greyscale images. The number of representations has today significantly increased with, for example, the Laplacian pyramid transform and the wavelet transforms, whether classical or geometrical. These new representations have led to considerable progress in applications such as compression, denoising and now characterization of the contents.

In this section, we show the way we followed to justify the choice of characterization approach based on geometrical wavelet transforms and especially on the Curvelets transform.

2.1 At the beginning, the Fourier transform

The initial work of Joseph Fourier focused on modelling the evolution of temperature through trigonometric series We call today those transforms: Fourier series. More precisely Fourier showed that for any periodic signal f of period T, we have:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{k=\infty} \left(a_k cos \left(\frac{2k\pi t}{T} \right) + b_k sin \left(\frac{2k\pi t}{T} \right) \right)$$

$$(1)$$

Coefficients a_k and b_k are given by:

$$a_k = \frac{2}{T} \int_0^T f(t) cos \left(\frac{2k\pi t}{T} \right) dt$$
$$b_k = \frac{2}{T} \int_0^T f(t) sin \left(\frac{2k\pi t}{T} \right) dt$$

$$(2)$$

Formula (1) may be rewritten with the module r_k ($r_k = \sqrt{a_k^2 + b_k^2}$) and phase θ_k ($\cos\theta_k = \frac{a_k}{r_k}$

and $\sin\theta_k = \frac{b_k}{r_k}$) :

$$f(t) = \sum_{k=1}^{k=\infty} r_k cos \left(\frac{2k\pi t}{T} - \theta_k \right)$$

$$(3)$$

The formulas (1) and (3) can be adapted to the case of non-periodic and integrable signals. This is called the Fourier transform.

The Fourier transform F is an operation that transforms an integrable function into another function, describing the frequency spectrum of f. If a function f is integrable, its Fourier transform is the function $F (f)$ and is given by the formula .

$$F(f) : s \rightarrow \hat{f}(s) = \int_{-\infty}^{+\infty} f(t) e^{-2i\pi st} dt$$

(4)

The departure set is the one of integrable functions f of one real variable t. The arrival set is the one of functions F (f) of one real variable s. Specifically when this transform is used in signal processing, we say that t is the time variable, f is in the time domain, that s is the frequency and F is in the frequency domain.

The Fourier transform is particularly well suited for stationary signals, so named because their properties are statistically invariant over time. If so, locating frequencies is meaningless.

The so-called natural images, especially document images, containing homogeneous zones with sharp transitions, generally located on the contours. This property makes them non-stationary signals, since the singularities appear in a non-homogenous way. The absence of localization of these singularities can be very penalizing. A solution may be found in the Gabor transform. The main well-known drawback of this approach is in the choice of the size of the sliding-window. Drawback solved with the wavelet theory.

2.2 The wavelet transform

In 1983, Morlet working on the analysis of seismic signals noted the inadequacy of the sliding-window Fourier transform due to the rigidity imposed by the fixed size of the window. He then decided to use a window of expanded or contracted size as needed. The idea of wavelets was born.

The wavelet transform uses basis functions to analyse and reconstruct a given signal. To understand this process, we shall now give a brief background on the properties of vector spaces for such reconstruction.

2.2.1 Vector and basis functions

A basis of a vector space V is a set of linearly independent vectors such that any vector v of V can be expressed as a single linear combination of these vectors in the database. There may be more of one basis for a given vector space each but each of them has the same number of vectors: this number is called the dimension of the vector space. For example, any vector of R^2 can be expressed as a linear combination of vectors (1,0) and (0,1).

Thus any vector is a linear combination of the basis of their vector space. This concept, expressed in terms of vectors, spread easily to functions by replacing the vectors of the basis with basis functions. Complex exponential functions are the basic functions for the Fourier transform.

If f (t) and g (t) are two functions of L^2 (all square integrable functions on a given interval), their scalar product can be calculated. This operation is used to calculate the wavelet transform as we recall in the following mathematical formalization:

Let f(t) a real function of real variable. Ψ is a function of zero average, centred around 0 and negligible outside a compact interval, called mother wavelet. The wavelet transform of f is:

$$g(a,b) = \frac{1}{\sqrt{a}} \int_{t=-\infty}^{t=+\infty} f(t) \Psi \left(\frac{t-b}{a} \right)$$

(5)

From the function Ψ, we can build a family of wavelets functions given by translations (coefficient b) and dilatations (coefficient a) of Ψ:

$$\psi_{a,b}(t) = \Psi\left(\frac{t-b}{a}\right)$$

(6)

where a is the frequency and b the time.

This definition shows that wavelet analysis is the similarity measure between the basis functions (wavelets) and the signal itself. Similarity is here understood as similar frequency content. The coefficients calculated in the transform indicate the degree of similarity between the signal and the wavelet at the current scale.

Unlike the Fourier transform, the mother function is not defined in the theory so that it is possible to choose the most appropriate one to the problem. However, apart from the orthogonality of the basis vectors to obtain a good reconstruction, the wavelet basis must meet two important properties to deserve the name : the admissibility and the regularity condition. We will not study those here but we will focus on what we are interested in : anisotropic structures.

2.2.2 To anisotropic structures

Within the great diversity of approaches to complement the properties of "classical" wavelets are a class of methods seeking to better characterize the anisotropic structures. These are the geometrical wavelets which can be adaptive or not.

Geometrical adaptive wavelets tries to adapt the best wavelet basis to the geometry of a given image. It becomes clear that whatever the approach chosen, a preliminary estimation of the geometry is required before decomposition. This estimation may however be done in several ways: edge detection, triangulation, regularity estimation … We may cite, as examples of this class of methods, the Bandelets (Le Pennec & Mallat, 2003), the oriented wavelets (Chappelier & Guillemot, 2005) or the Directionlets (Velisavljevic, 2005) It appears that with such an approach, there is near no wavelets coefficients redundancy which can be an advantage for applications like compression of images.

Apart from characterizing the contours or local singularities, typically anisotropic properties, the non-adaptive geometrical wavelets share the distinction of having a fixed basis and independent of the image they represent. This allows, among other things, of not requiring additional cost for specifying, during the synthesis, the configuration used in the analysis. However this leads to many unwanted redundancy for applications such as compression. We may cite, also as examples of this class of methods, the Ridgelets (Candès & Donoho, 1999a), the Curvelets (Candès & Donoho, 1999b) and the Contourlets (Do & Vetterli, 2005)

2.2.3 The Curvelets

Due to fact that we are working with anisotropic structures, writings, we have chosen to use geometrical wavelets. The choice of a non-adaptive method is due to that we do not search to make compression of our signal but more to analyse the latter. Indeed, analysis may be better with a little redundancy in wavelets coefficients than if there was not. Finally the choice of Curvelets is due to that it has an optimal coding of information. The question now is how it is done. In the discrete domain and particularly in the case of images, we may

consider that locally, there are straight contours. This is what leads to the creation of the Curvelets transform. This transform is achieved in two main stages. First we partition the image into squares of varying sizes with recovery to avoid sides effects. These squares are obtained through a finite support Fourier window. The next step is to apply a discrete Ridgelets transform within these squares with dilation of the wave function of a/a^2. Contours not captured by the separable wavelet analysis can be found in the sub-band detail. A sufficiently fine partition of the sub-bands can then obtain blocks where these contours are straight lines and are therefore suitable for Ridgelets analysis. The Curvelets transform is invertible but redundant because the analysis in discrete sub-Ridgelets core is made using a Fast Fourier Transform (FFT) of the polar plane, requiring more points than those available in the rectangular grid. Using the FFT directly from the Fourier Slice Theorem. Indeed, it indicates that the Radon transform can be obtained by applying an inverse 1-D Fourier transform along radial lines passing through the origin in the Fourier domain of the 2-D image. All these steps are illustrated in Fig. 2.
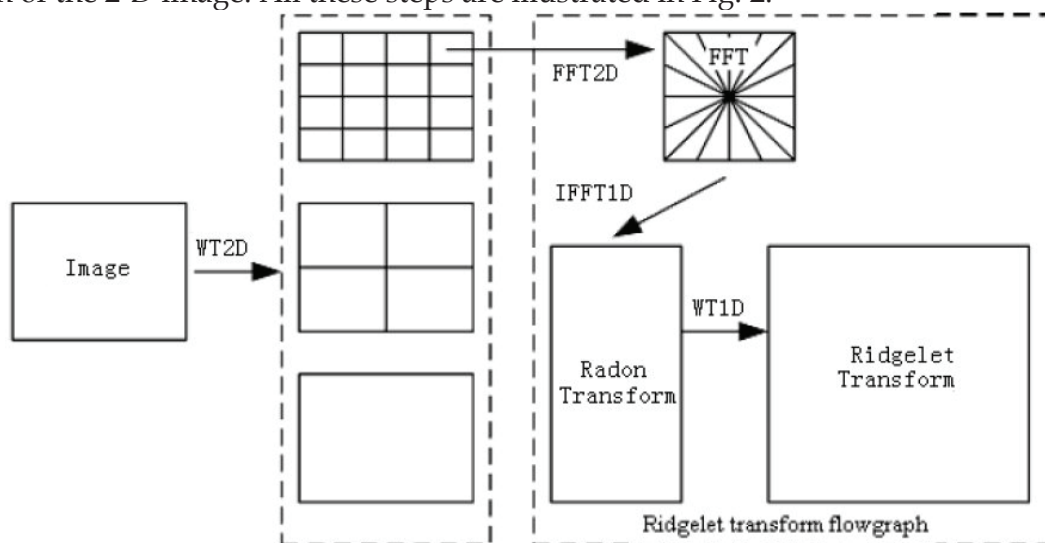

Fig. 2. Curvelets transform flow graph on an image

3. Features extraction

3.1 DCTG2

The version of the Curvelets transform introduced in section 0 uses the Radon transform. This transform is difficult to discretize. Candès and Donoho have proposed a second generation of transform that are now developed in the discrete domain. The development has been done in the software Curvelab (http://curvelet.org) The technical details related to this implementation are provided in (Candès et al, 2005) We have chosen to use in our work this implementation for the extraction of orientations and the evaluation of the curvatures on each point of the transformed image.

3.2 Orientations extraction

Curvelets coefficients are indexed in position, scale and direction. The coefficients enabled us to deduce easily the dominant orientations of strokes by retaining only the highest coefficient corresponding to processed pixel at the finest scale. We only retain the finest

scale for the analysis of orientation and curvature. However, the methodology presented can be applied regardless of the scale.

In theory, for a given scale E, the number n of sub-bands created by the directional Curvelet transform is calculated using the formula :

$$n = a * 2^{\lfloor \frac{(E-1)}{2} \rfloor}$$

where a is the number of angles used for the first cutting angular frequency plane which is in our case the scale 2.

In practice we set the number of scales and a to 8 and therefore we obtain at the finest scale 64 angles: so the accuracy of the directional decomposition is around 5°. Searching for a greater accuracy requires an additional cost in terms of storage that is not necessary for our work. In addition, this implies sensitivity to slight variations in the writing which is not desired for the great variability induced by production of handwriting.

Thus, if Op is the set of coefficients of the n orientations that are analysed at the finest scale corresponding to pixel p, then the dominant orientation is o with $o = index\ (MaxOp)$ where *Index (j)* is a function returning index $i <= n$ corresponding to Curvelet coefficient j.

We show on Fig. 3. the results of this analysis on an extract from the Washington's manuscripts library. Analysis of the circle on Fig. 4. is given as a colour reference and a zoom on a part of the orientation analysis is given on Fig. 5.
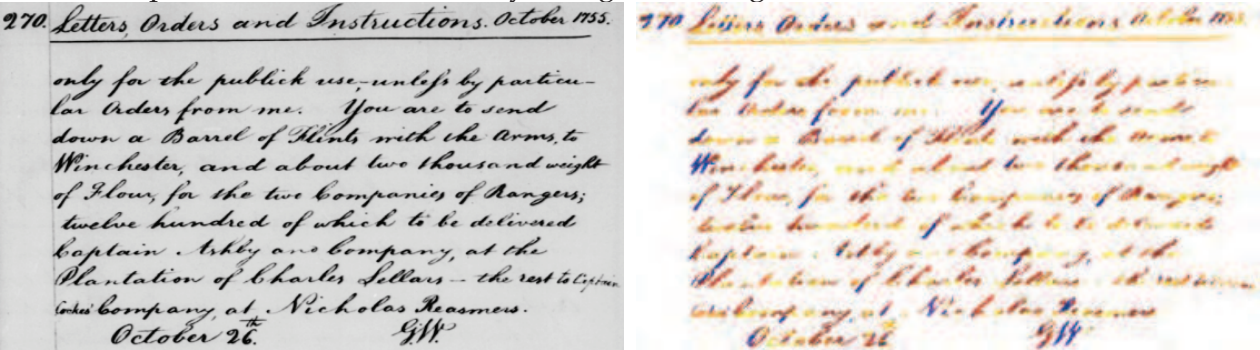

Fig. 3. Orientation analysis based on the Curvelets transform

Fig. 4. Orientation analysis of a circle

Fig. 5. Zoom on a part of the orientation analysis

3.3 Curvatures evaluation

We just show how we extract the dominant orientations. The latter, although adequate for right segments, had to be enriched for all other cases, especially for handwriting that we study but also in almost cases of images in general. This enrichment has come to the attention of the other dominant orientations for the same pixel. Indeed, it is rare for a pixel of a curve (or contour) to have a single dominant orientation and it is very common that other orientations are also carriers of information. Therefore, for each pixel, we keep a list of different dominant orientations associated with it. This list of dominant orientations has been inspired from Antoine and Jacques's works in (Antoine & Jacques ,2003)

In their work, authors use conical wavelets (Antoine et al, 2008). The support of these wavelets is strictly in a convex cone in the frequency domain which makes them very efficient in detecting objects whose orientation is marked as straight lines. From these wavelets, authors have shown that it is possible to estimate the curvature at a point in a line by estimating the number of orientations for which the wavelet coefficients corresponding to this point are significant. The main problem in our work by using this approach comes from the limitation to straight lines. Considering that conical wavelets approximate heavily Curvelet except that the parabolic scaling that is specific to Curvelets transforms allows a better selectivity of anisotropic objects. In that context, we decided to apply the same methodology but this time with the Curvelets transform. In that way we have used the list of dominant orientations that we have mentioned earlier to evaluate the curvature at the frontiers of lines and shapes. The evaluation of the curvature on each point is then simply computed by measuring the length of this list. The more long it is, the more the curvature at the point of interest is high and vice versa. Formally, we can define our evaluation of the curvature as follows:

Let's consider P as a pixel of an image I. Lp represent all significant orientations associated with it, then the level of curvature in P, Np is defined by:

$$Np = Card\ (Lp)$$

A result of this evaluation is given on Fig. 6. for the same document as for the Fig. 3. Colours used for this representation are ordered on Fig. 7. and a zoom on a part of the evaluation is given on Fig. 8.
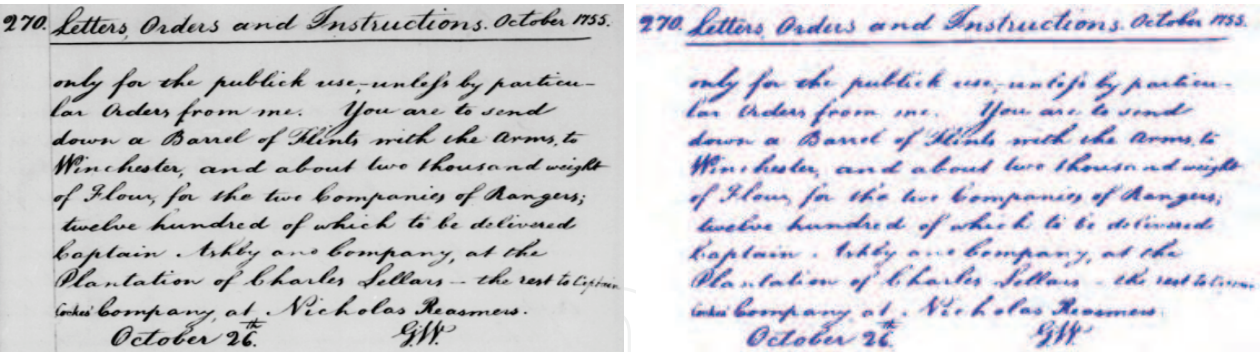
Fig. 6. Curvature evaluation based on the Curvelets transform

Fig. 7. Standard curvature scale

Fig. 8. Zoom on a part of the curvature evaluation

3.4 Handwriting signature design

The analysis of orientations and curvatures at the frontiers of shapes that we have presented in sections 0 and 0 provides a characterization of writing in the manuscript. However, as the amount of information contained in a manuscript may considerably vary, we have built a compact representation of this information which allows to describe a unique writing contained in the document. This representation is what we call the signature of writing and is the score matrix M of couples (curvature, orientation) extracted from the analysis of orientations and curvatures presented in sections 0 and 0.

Thus, for each pixel p on frontiers of writing, we build the couple (c , w) where c is the curvature of the line in pixel p and w is the set of c orientations retained significant for the evaluation of the curvature. We raise the values of matrix M with coordinates (c , o_i) of a value corresponding to the ratio between the coefficient o_i and the maximum of coefficients all o_i considered (see Fig. 9).
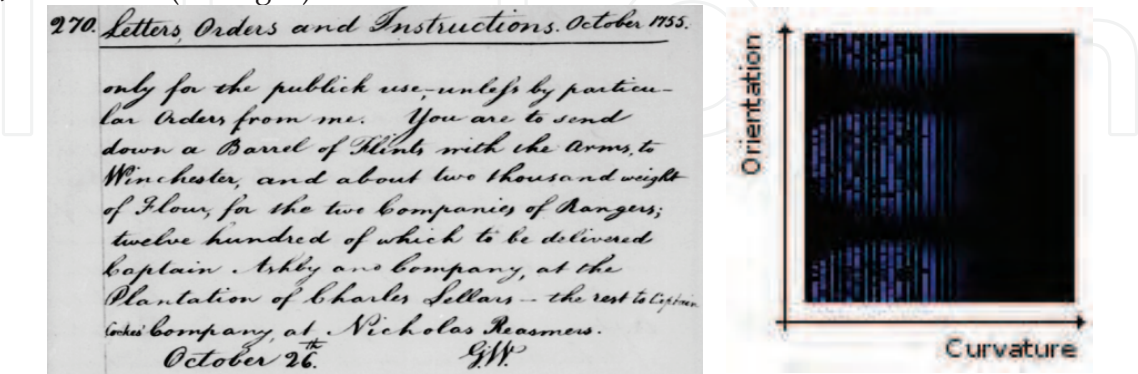

Fig. 9. Signature of Washington's handwriting

Values of this matrix of occurrences remained closely linked to the amount of information contained in the manuscript so we standardized values by keeping the ratio between the

maximum value of the signature and the values of the latter. The values of the signature are then all between 0 and 1.

At the same time we have removed from our signature odd columns of the matrix. Indeed, the analysis takes into account the 360 degrees and therefore evaluates each direction twice (45 ° and 225 ° for example). This produces a curvature systematically even hence our choice (see Fig. 10). We have also removed half of the lines, but in order to illustrate and justify the symmetry, we will still represent them in the following.
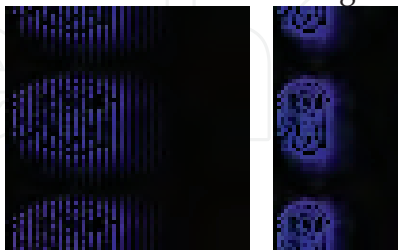
Fig. 10. Even curvatures suppression in signature

We quickly found that the standardization is not suitable for the comparisons intended for this signature. Indeed, with such an approach, significant values of the signature are, in the same language, almost invariant. For the Latin languages, for example, there are almost systematically the vertical and / or horizontal orientation considered as dominant and thus around a curvature evaluation rather low. This comes, of course, essentially from the verticality of our scripts and the horizontal alignment of our texts. To solve this problem, we have raised low values of our signature while reducing the role of orientations common to all writers. To do so, we applied a Lorentzian filter (see Fig. 11 { 2 }) before reversing the order of values in the signature (see Fig. 11 { 3 }):

$$\forall i, \forall j \rightarrow M(i,j) = 1 - L(M(i,j))$$

where L is the Lorentzian filter defined as :

$$L(x) = \frac{\Gamma}{2\pi} \frac{1}{\left(\frac{1}{2}\Gamma\right)^2 + (x - x_0)^2}$$

where $x0$ is the index of the peak and Γ its width at half height.
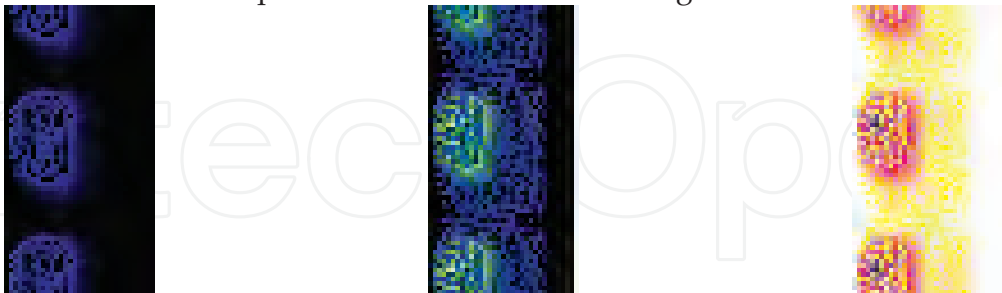
{ 1 } Signature after even curvatures suppression

{ 2 } Signature after applying Lorentzian filter

{ 3 } Signature after reversing order of values

Fig. 11. Different steps of our signature changes

4. Applications

Now that features can be extracted and that if necessary those can be grouped together in a compact signature, we will now show how we used them to retrieve information from

manuscripts. In section 0 we will present how we have built a CBIR (Content Based Image Retrieval) system based on our compact signature of handwritings and in section 0 we will present a word spotting system based, for the moment, exclusively on the orientations extraction.

4.1 Content based image retrieval

As every CBIR system, our one is composed of two main parts: a features extraction phase that allows a change in representation space and a comparison phase between those features which provides a list of similar images. This scheme of operation is identical to the one of a k nearest neighbours approach and is presented in Fig. 12.
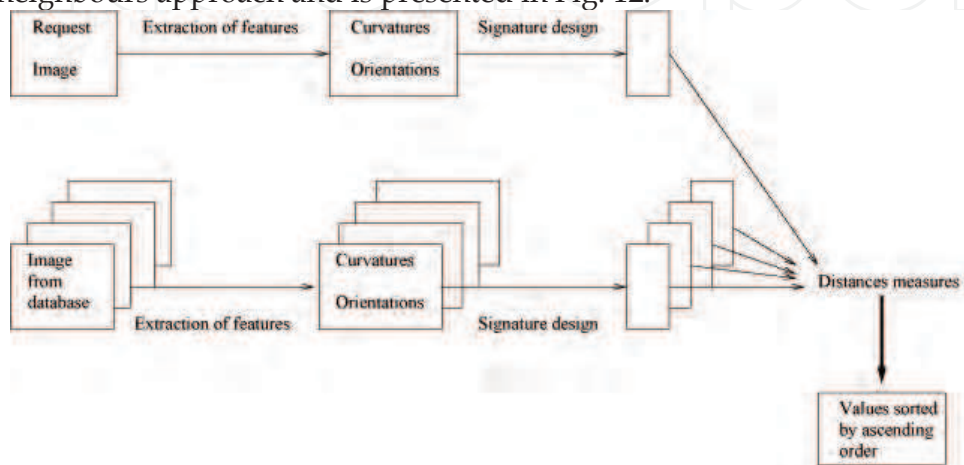
Fig. 12. Schema of a content based image retrieval system

The features extraction and the signature design steps have been presented in part 0 and we will now have a look on the similarity evaluation between signatures.

We have developed two main measures; a generic one and another one dedicated to medieval Latin manuscripts. The two measures have been integrated in our system and we have tested those on two different databases.

The evaluation process was the same for both:

_ The precision P is evaluated over all queries: $P = \dfrac{NumberOfGood\,Answers}{NumberOfAnswers}$

_ The recall R is evaluated over all queries: $R = \dfrac{NumberOfGoodAnswers}{NumberOfExpectedAnswers}$

_ The F-measure is computed from P and R: $F = \dfrac{2PR}{P+R}$

To highlight those values we have decided to compare our system to another one from the state of the art. Many of those use features based on texture, colour, … and in our case comparison would not be very easy. So we decided to test one based on wavelets coefficients: ImgSeek (Jacobs et al, 1995)

4.1.1 Results of the generic measure

The generic measure is a linear correlation coefficient and we have tested it on a humanistic handwritings database. We present here the results obtained on the basis of manuscripts

IAM. The value of this database is that information about writers are known and the types of scripts are closed to what we call humanistic writings. We also tested our approach on images of ancient manuscripts with similar types to those in the IAM database and results were similar. However we do not present these results for rights issues on these documents. As we have mentioned, all images of the IAM database are given with accurate information about writers, the task of labelling was greatly facilitated. We provide results in Table 1 and the corresponding curves in Fig. 13.

System	Precision	Recall	F-measure
Our system	0,87	0,68	0,81
ImgSeek	0,82	0,63	0,73

Table 1. Comparison between characteristic values of our system and ImgSeek on Humanistic database
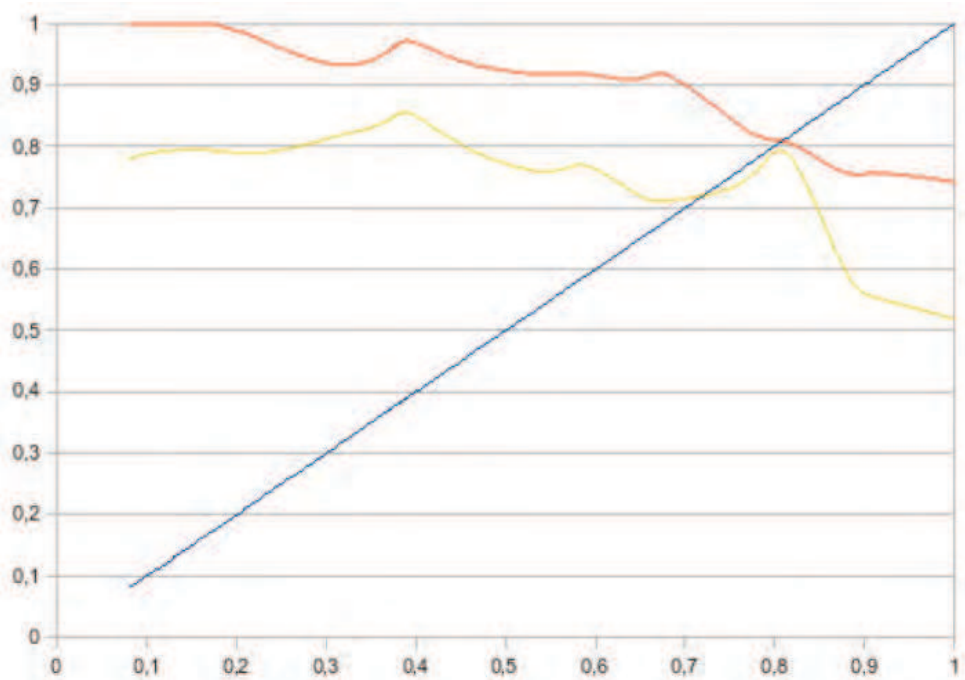
Fig. 13. Precision/Recall curves of CBIR on Humanistic database. Our system in red and ImgSeek in yellow

It appears that, despite our similarity measure do not take into account information on handwritings properties, results stay better with our system. We have also tested our approach on a natural images database and results are almost the same except that gap between the two systems is greater.

4.1.2 Specialized similarity index

We have the chance to work with palaeographers and this led us to study our linear correlation on their images. Results were clear not what they expected. By the way, we have tried to understand where could be the problem of our approach and it appears clearly that the linear correlation take to much into account some part that should not. Two Latin handwritings have clearly some common properties that should be considered in the same

way that their differences. This assessment, or nearly the same, has been done by Tversky (Tversky, 1977) when he built is ratio model in the cognitive psychology field. The problem for us of this model is that it is based on binary features which is clearly not our case. We have then decided to adapt it to our specific signature. To do so, we decided to separate information of our signatures on the basis of their common and what we call the residual or distinctive parts. We shall now see how we make this separation.

For separation of values into two groups, we defined the combination of two measures that are:

_ A linear correlation based on common parts of the two signatures, which we denote $Cor_{directe}$.

_ A linear correlation based on opposition parts between the two signatures, which we denote $Cor_{residus}$.

To define common parts between the two signatures we set a threshold S between the strong and low values. This threshold is the mean of the first signature that is in our case the query signature. If for a couple of values in the signature, values of both signatures are higher (respectively lower) to S then they would be taken into account in the calculation of a direct correlation denoted Cor_{forte} (Cor_{faible} respectively) (see Fig. 14)

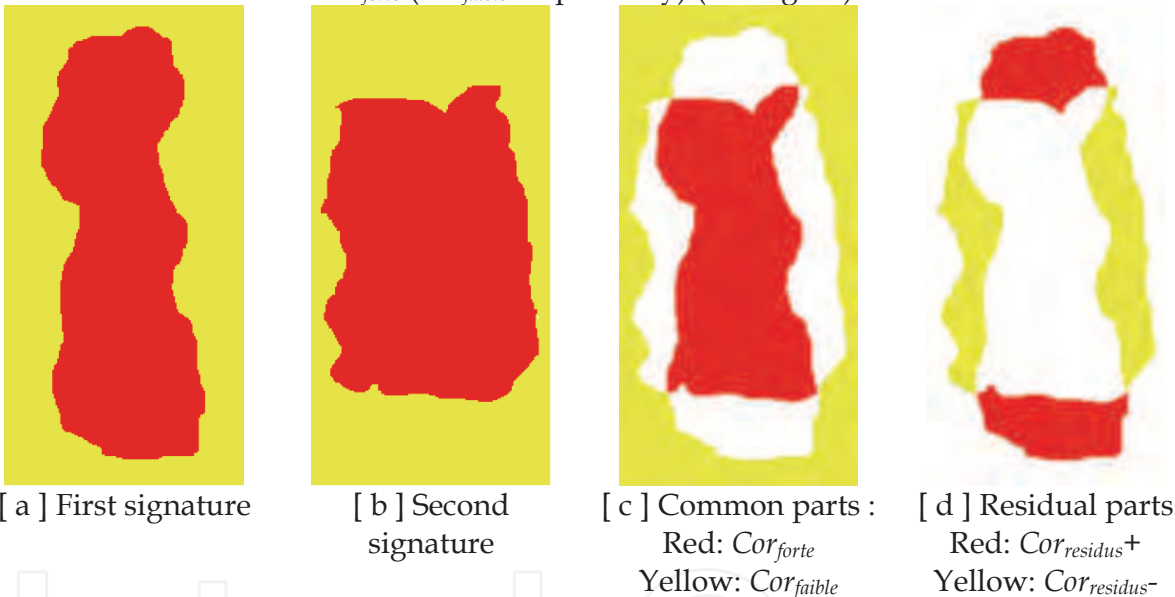| [a] First signature | [b] Second signature | [c] Common parts : Red: Cor_{forte} Yellow: Cor_{faible} | [d] Residual parts Red: $Cor_{residus}+$ Yellow: $Cor_{residus}-$ |

Fig. 14. Parts kept for the direct correlation and the residual correlation

Cor_{forte} and Cor_{faible} are grouped together in a single measure defined as follows:

$$Cor_{directe} = \alpha \times Cor_{forte} + (1 - \alpha) \times Cor_{faible} \qquad (7)$$

α depends on number of values kept in each part.

In exactly the same way $Cor_{residus}+$ and $Cor_{residus}-$ are grouped together in another single measure defined as follows:

$$Cor_{residus} = \beta \times Cor_{residus+} + (1 - \beta) \times Cor_{residus-} \qquad (8)$$

β also depends on number of values kept in each part.

Our index is then defined as :

$$Sim = \gamma \times Cor_{directe} + (1 - \gamma) \times Cor_{residus} \qquad (9)$$

This time, the value of γ is left to the user. He may then choose between taking more into account the common parts of signatures or what makes the difference between them. On our

medieval manuscripts database, curves of Precision/Recall according to the value of γ are given in Fig. 15.
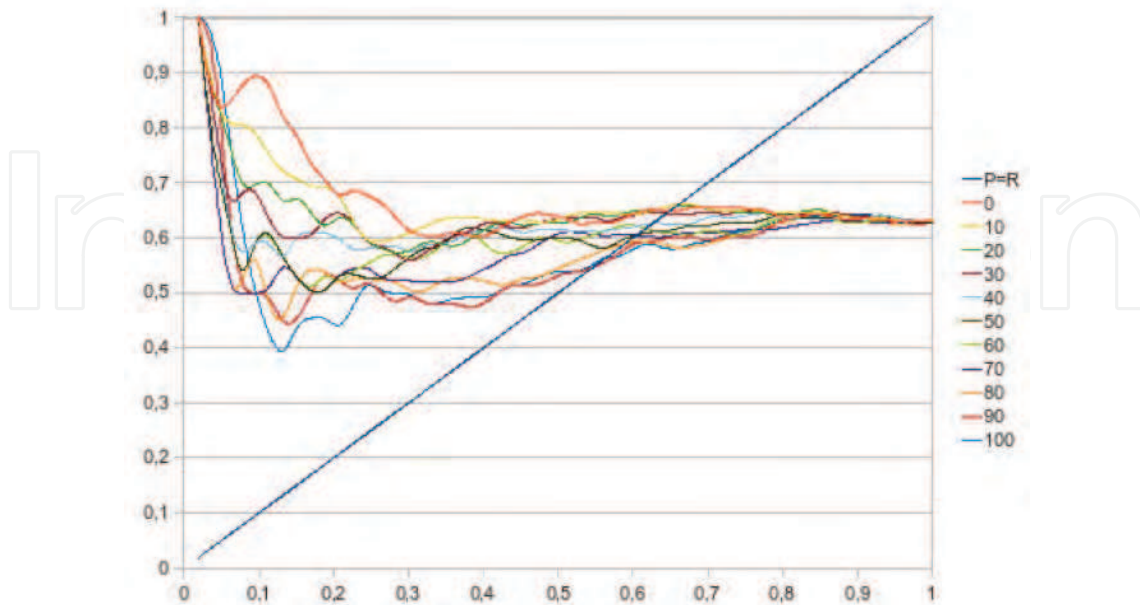

Fig. 15. Our system behaviour according to values of γ

What we can see on this curves is that, for this specific database, it is clearly more interesting to consider differences between writings more than common parts. Indeed, despite the end of curves is almost the same, for low values of recall (first answers) the precision stay higher with low values of γ

4.2 Word spotting

The word spotting is based on a similarity or a distance between two images, the reference image defined by the user and the target images representing the rest of the page or all the pages from a multi-page document. Firsts results in handwriting documents have been given by Manmatha et al. in (Manmatha & Croft, 1997). Their idea was to match words from a request and words in documents by the use of simple features such as aspect ratio's of word's bounding box. Other approaches use direct correlation methods applied on the grey levels for image similarity comparison. Those classical approaches are very sensitive to geometrical transformations and are time consuming. Moreover, correlation cannot be easily adapted to the spatial variations of the handwriting. Main solutions consist in representing the informative parts of the images with feature vectors that can be compared with other feature vectors from other images. The choice of the most discriminant features to characterize regions of interest is central. Different approaches for local image structure description have been proposed in the literature. We propose here a short overview of those methodologies. An approach proposed by Kolcz et al. in (Kolcz et al, 2000) was to search along lines of text a way to match their request in every position along this line. Several features have been used and are matched with a dynamic time warping (DTW) distance. This is a expensive way to search a match but some heuristics are used to limit the search along the lines. Moreover, results are really much better than those proposed in (Manmatha & Croft, 1997). A mix of this kind of ideas can be found in (Rath & Manmatha, 2003). The

idea is to segment each documents into words with a technique given in (Manmatha & Srimal, 1999), and then to resize and realign each word so that it can be easily compared with other images. Distance between images is then computed with the DTW distance on a set of features described in this paper. The critical part of this approach is in the segmentation into words which can not be done on every documents and especially on medieval documents.

Recently, Leydier et al. in (Leydier et al., 2007) have proposed a (word, line, layout) segmentation-free method which rests on the assumption that words can be matched on a small number of guidelines. Very good results are shown but no idea is given on what happened if text is not as vertical and straight as those in manuscripts studied. However, they have proved that the strokes' orientation provide a strong piece of information in handwritten manuscripts. The hitch comes from matching distances tested which are based on the assumption that guidelines are vertical which is not necessary the case in other manuscripts. Other keyword spotting approaches very close to our proposition have been proposed those last years. Fink and Plotz in (Fink & Plotz, 2005) have tested appearance-based features for writer independent handwritten text recognition and compared it with heuristic features. Terasawa et al. in (Terasawa et al., 2006) have developed principal component analysis-based descriptors and gradient distribution features for word spotting in historical handwritten documents. The only limitation of the approach is its application to only well segmented threshold documents and very regular handwritten texts.

Our approach tries to keep the best of the different approaches presented, by searching guidelines but not necessary vertical ones. Our idea is to search for the predominant orientations of the query in documents (see Fig. 16).

[a] Document image   [b] Orientations extraction   [c] Query's main orientation

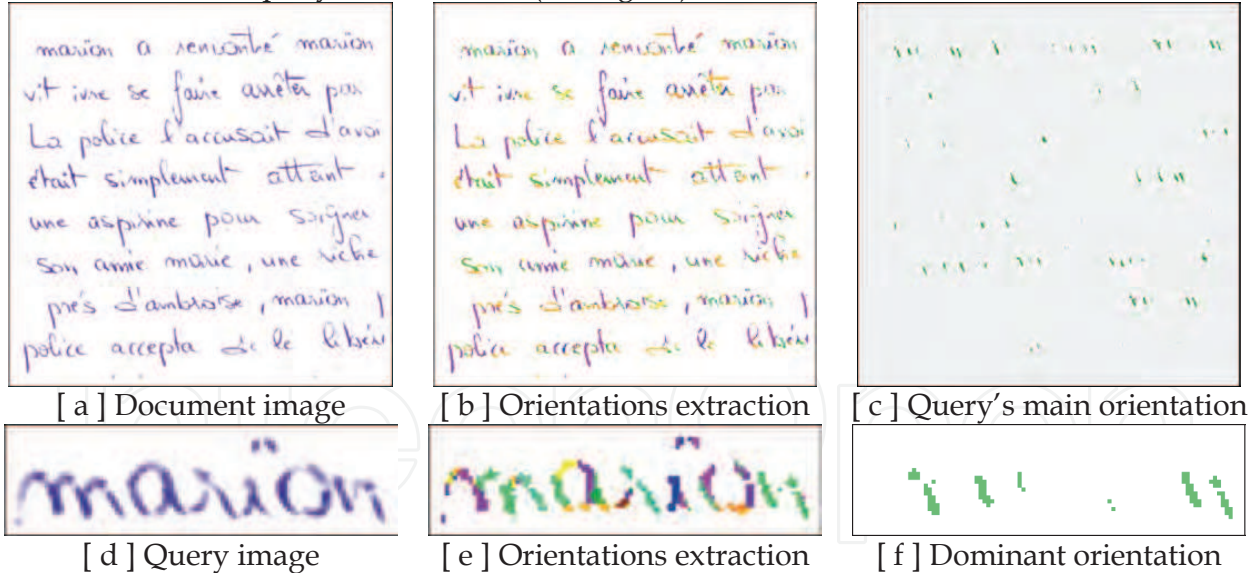[d] Query image   [e] Orientations extraction   [f] Dominant orientation

Fig. 16. Extraction of the dominant orientation of a query

Once orientations are extracted from the request by the use of the Curvelet transform, we search for the same organization of those orientations in a window similar to the one of the request sliding over the documents. The organization is obtained from a quad tree applied on the window in which each leaf contains the area covered by exactly the same number of pixels detected with predominant orientation of the query image (see Fig. 17).

[a] Dominant orientation of the query

[b] First level of the quad tree

[c] Second level   [d] Second level   [e] Second level   [f] Second level

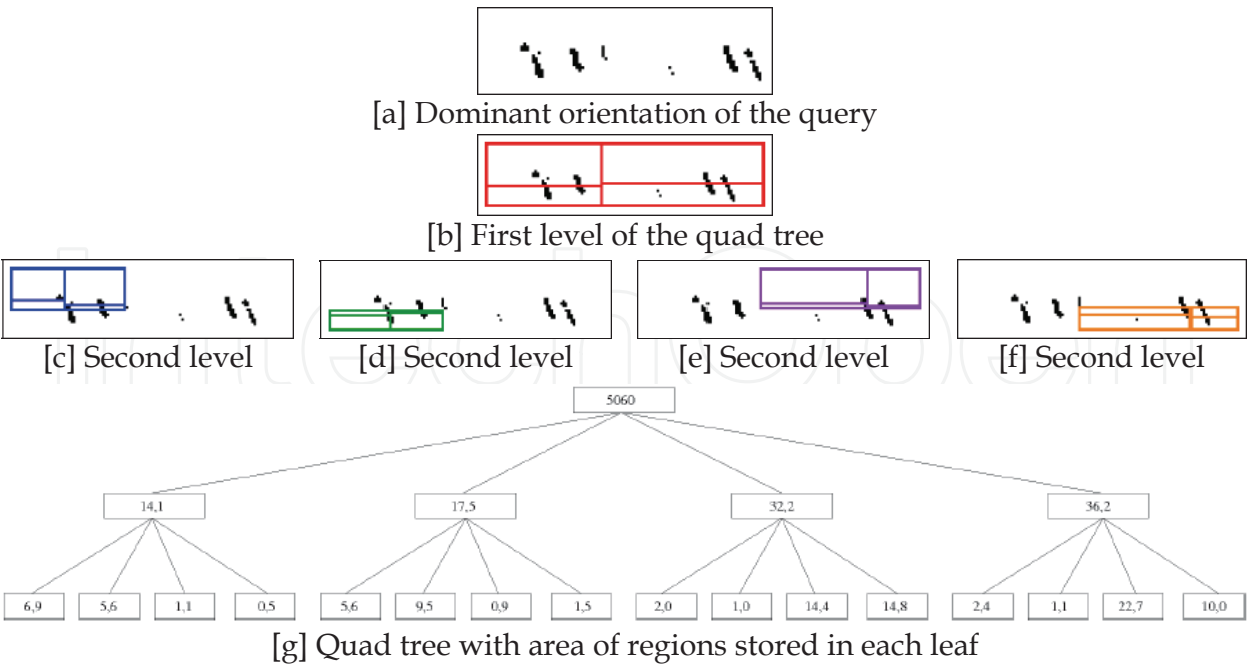[g] Quad tree with area of regions stored in each leaf

Fig. 17. Dominant orientation pixels organization evaluation with a quad tree

As it is also an information retrieval system, we have computed the same value that what we have done in section 0. This time we have compared our system to the one of Leydier et al (Leydier et al., 2007) This time, it is our system which use less information than the one we compare to. Indeed, the system of Leydier et al uses information on what it searches for. For example, guidelines used to search for a word are vertical which is due to fact that this system was initially developed for Latin writings. Nevertheless, it is the closer to our system and that is why we decided to compare our system to this one.

The first database we tested is the Washington's manuscripts database. Characteristic values of this comparison are given in Table 2 and corresponding curves are given in Fig. 18.

System	Precision	Recall	F-Measure
Our system	0,8	0,42	0,55
Leydier et al.	1,0	0,53	0,69

Table 2. Comparison between characteristic values of our system and Leydier et al. on Washington's manuscript database
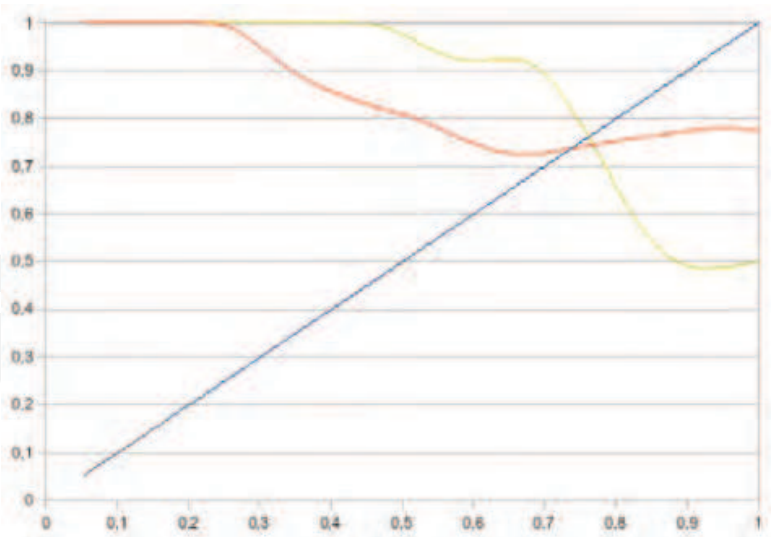
Fig. 18. Precision/Recall curves of word-spotting on Washington's manuscript database. Our system in red and Leydier et al. in yellow

Leydier et al. system is better than our one because of the use of information on shape searched. If we work on a database of comics images, which are also strokes images, the result are almoste inverted due to the generic aspect of our approach. Characteristic values are given in Table 3 and corresponding curves in Fig. 19.

System	Precision	Recall	F-Measure
Our system	1,0	0,4	0,57
Leydier et al.	0,9	0,36	0,51

Table 3. Comparison between characteristic values of our system and Leydier et al. on comics images database
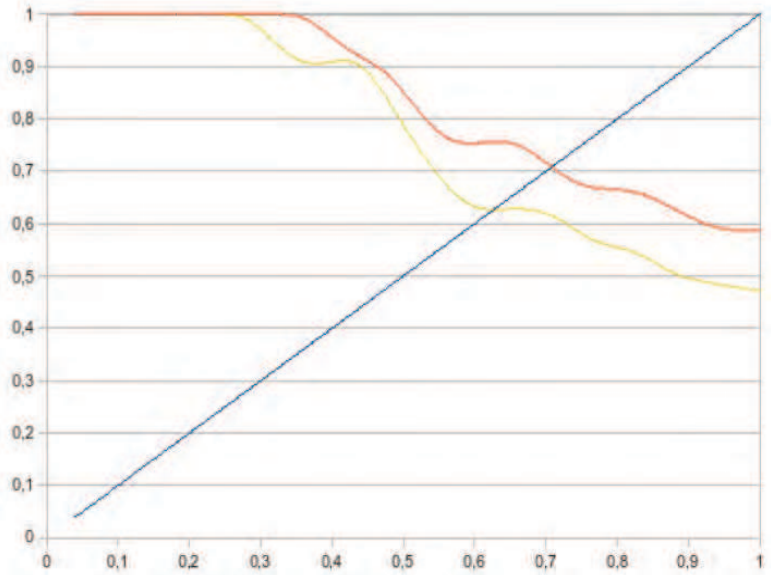
Fig. 19. Precision/Recall curves of word-spotting on comics images database. Our system in red and Leydier et al. in yellow

5. Conclusion

We have shown how one can extract features, curvatures and orientations in frontiers of shapes, from a geometrical wavelets transform. We have also shown how we can use these features to develop two applications of information retrieval. The content based information retrieval system is based on a global information about the handwriting in the document and the word spotting system is based on local information about a keyword. We have compared those systems to the state of the art and we are currently working on how to improve results.

For the word spotting system, we consider that we can use the curvature information to locate interesting points in the documents. Currently, we slide our window over all the document pixel by pixel. This is very expensive and as we can see on Fig. 20 if we reduce search to, for example, high curvatures zones the treatment cost should be greatly reduced.

[a] A document image   [b] Curvature evaluation   [c] Curvature thresholded
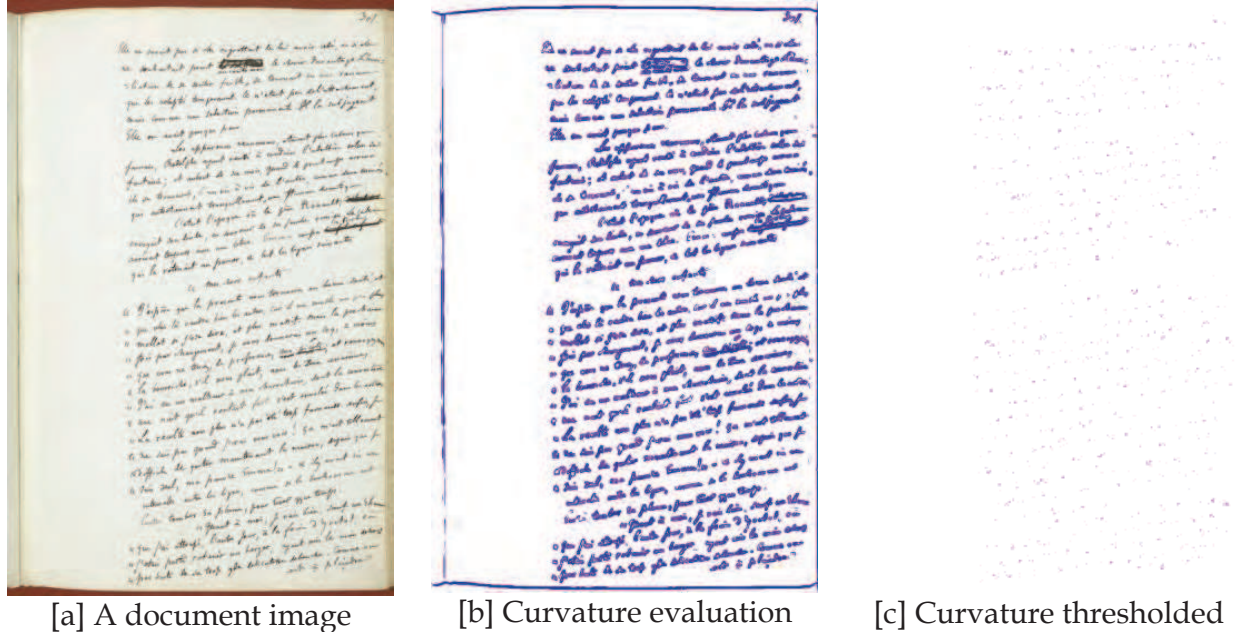
Fig. 20. Example of reduction of search field thanks to curvature

We are also working on an improvement of our CBIR system. This improvement comes from a density study of writings in a new representation space. This study is closed to the Parzen window method. Once it is done, we may use this study as a preliminary index to the search in the database. We are currently working on this approach on our medieval manuscripts images database. Results still need to be evaluated by experts before we can apply it in our CBIR system.

6. References

Antoine, J-P & Jacques, L. (2003). Measuring a curvature radius with directional wavelets. Institute of Physics Conference Series, (899-904) ,(2003)

Antoine, J-P; Bogdanova, I. & Vandergheynst, P.(2008). The continuous wavelet transform on conic sections. International Journal of Wavelets, Multiresolution and Information Processing, 6,2,(137-156) ,(2008)

Candès, E. & Donoho, D., (1999). Ridgelets : a key to higher-dimensional intermittency ?, Philosophical transactions - Royal Society. Mathematical, physical and engineering sciences, 357, (2495-2509) ,(1999)

Candès, E. & Donoho, D., (1999). Curvelets : A surprisingly effective non-adaptive representation of objects with edges. Technical report

Candès, E.; Demanet, L.; Donoho, D. & Ying, L. (2005). Fast Discrete Curvelet Transforms., Technical report

Chappelier, V. & Guillemot, C.(2005). Oriented wavelet transform on a quincunx pyramid for image compression. IEEE International Conference on Image Processing, 1:I, (8-14) ,(2005)

Do, M.N. & Vetterli, M., (2005). The contourlet transform : an efficient directional multiresolution image representation. IEEE Transactions on Image Processing,14,12,(2091-2106) ,(2005)

Fink, G. A. & Plotz, T., (2005). On appearance-based feature extraction methods for writer-independent handwritten text recognition. International Conference on Document Analysis and Recognition, (1070-1074), (2005)

Jacobs, C.E.; Finkelstein, A. & Salesin, D.H.(1995). Fast multiresolution image querying. ACM, (277-286),(1995)

Kolcz, A.; Alspector, J.; Augusteijn, M.; Carlson, R. & Popescu, G., (2000). A line-oriented approach to word spotting in handwritten documents. PAA, 3,2 (153-168), (2000)

Le Pennec, E. & Mallat, S. (2003). Bandelettes et représentation géométrique des images. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images

Leydier, Y. ; Lebourgeois, F & Emptoz, H., (2007). Text Search for Medieval Manuscript Images. Pattern Recognition, 12,40, (3552-3567), (2007)

Manmatha, R. & Croft, W., (1997). Word spotting: Indexing handwritten archives.Intelligent Multimedia Information Retrieval Collection.

Manmatha, R. & Srimal, N. (1999). Scale space technique for word segmentation in handwritten manuscripts. ScaleSpace99, (22-33), (1999)

Rath, T. M. & Manmatha, R., (2003). Features for word spotting in historical manuscripts. International Conference on Document Analysis and Recognition,(218), (2003)

Terasawa, K. ; Nagasaki, T. & Kawashima, T. (2006). Automatic keyword extraction from historical document images. Document Analysis Systems, (413-424), (2006)

Tversky, A., (1977). Features of similarity. Psychological Review, 84,4,(327-352), (1977)

Velisavljevic, V., (2005). Directionlets. PhD thesis, Lausanne
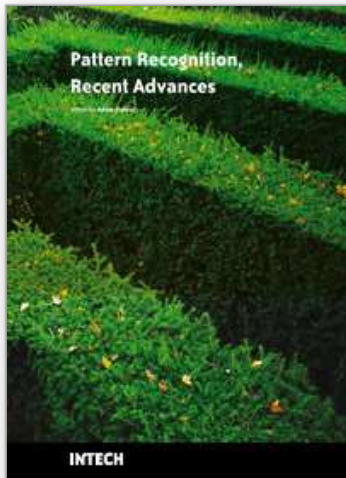
Pattern Recognition Recent Advances

Edited by Adam Herout

Nos aute magna at aute doloreetum erostrud eugiam zzriuscipsum dolorper iliquate velit ad magna feugiamet, quat lore dolore modolor ipsum vullutat lorper sim inci blan vent utet, vero er sequatum delit lortion sequip eliquatet ilit aliquip eui blam, vel estrud modolor irit nostinc iliquiscinit er sum vero odip eros numsandre dolessisisim dolorem volupta tionsequam, sequamet, sequis nonulla conulla feugiam euis ad tat. Igna feugiam et ametuercil enim dolore commy numsandiam, sed te con hendit iuscidunt wis nonse volenis molorer suscip er illan essit ea feugue do dunt utetum vercili quamcon ver sequat utem zzriure modiat. Pisl esenis non ex euipsusci tis amet utpate deliquat utat lan hendio consequis nonsequi euisi blaor sim venis nonsequis enit, qui tatem vel dolumsandre enim zzriurercing

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds