

# Generic scale-space process for handwriting documents analysis

Guillaume Joutel, Véronique Églin, Hubert Emptoz  
LIRIS UMR CNRS 5205 - INSA Lyon, 69621 VILLEURBANNE Cedex  
{guillaume.joutel;veronique.eglin}@insa-lyon.fr

## Abstract

*This paper presents a generic architecture for handwriting documents analysis. It covers all analysis steps from the content description of the document (layout analysis, handwriting shape characterization) to three dedicated Digital Libraries applications (CBIR<sup>1</sup> in great ancient documents images database, Paleographical images classification and word spotting). The generic scale space tool is based on the Curvelets decomposition of images for the indexation of linear singularities of handwritten shapes. 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 use the Curvelets coefficients in various ways to satisfy the three designed applications. The complete implementation scheme is validated with a specific application of word spotting based on the orientations analysis. The proposed method is language independent and only visual orientation and appearance based. In that context, no lexical information nor any other statistical language models are required. The first proposed tests for this application are proposed on medieval documents images and on European 18th century correspondences corpus from the CER-PHI. Precision-recall analysis testifies the relevance of the contribution.*

## 1. Introduction

In this work, we are interested in ancient handwriting documents images analysis (digitized Middle-Age composed by copyists' texts from the 9<sup>th</sup> to the 15<sup>th</sup> century and Humanistic manuscripts essentially composed by authors' drafts from the 18<sup>th</sup> and 19<sup>th</sup> century). We intend to propose generic and dedicated tools for palaeographers and historians so as to help them in their expert

work of manuscripts dating, categorization and authentication.

Primary difficulties faced by this community of experts are the classification of handwriting documents images, the hands identification in specific handwritten collections and the information and content retrieval in great images database (by layout similarity and by keyword or shapes indexing). In this paper, we propose a contribution to this challenge that must be considered as a tool kit for historians the most generic as possible. We have focused on a generic methodology that exploits a scale space theory based on Curvelets transform introduced by Candès and Dohono for signal processing. We have developed it in different ways to characterize, classify, retrieve and index handwriting documents images from those ancient corpus.

Due to the fact that historians experts<sup>2</sup> judge the *curvature* and the *orientation* as two fundamental dimensions of handwritings, we have searched a way to compute them on the handwritten shapes and to use them as visual features for the characterization, the classification, and the retrieval of documents images for different handwritings collections. To compute those dimensions, we have developed a methodology that is sensitive to variations at the frontiers of 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.

In this paper, we will focus on a specific application: the word spotting that consists in localizing users' selected words in a handwritten document without any syntactic or lexical constraint. This technique is used when word recognition cannot be done, for example on very deteriorated printed documents or on manuscripts. We have developed a generic approach that can be applied to any written document in any language using any al-

<sup>1</sup>Content Based Image Retrieval

<sup>2</sup>Pr. D. Muzerelle and Pr. D. Poirel from the IRHT in Paris

phabet, pictograph or ideogram. As a result, the system proposes a sorted list of hits that the user can prune manually.

## 2. The generic system based on a Curvelets decomposition

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 located 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. One solution to this problem is the Curvelet transform. It is more robust than wavelets for the representation of shapes anisotropy, of lines segments and curves in the images. 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:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, r \in [3/4, 3/2]$$

$$\sum_{t=-\infty}^{\infty} V^2(t - l) = 1, t \in [-1/2, 1/2]$$

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. 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, [1].

Our generic system is based on this geometrical multiscale images decomposition embedded in its different stages. The figure 2 presents the system architecture with the first part corresponding to the analysis of the document and the second part corresponding to the three main applications currently under studies.

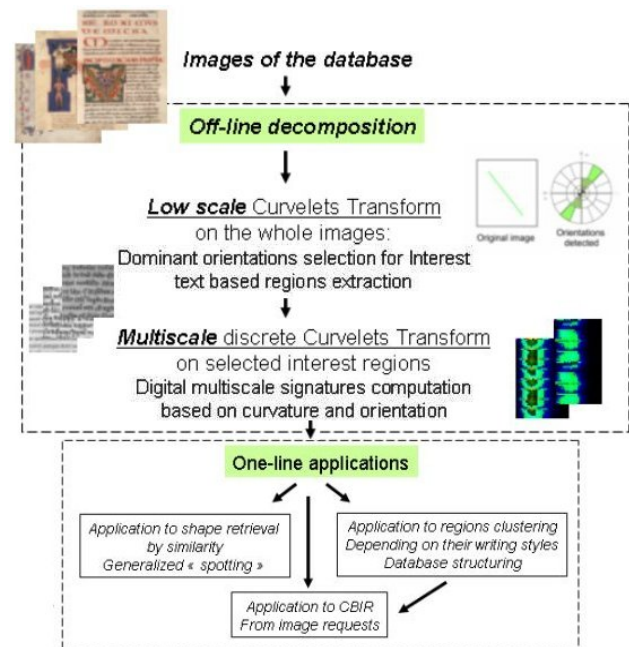


Figure 1. Architecture for the handwriting characterization and on-line applications

## 3. 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. First results in handwriting documents have been given by Manmatha et al. in [6]. 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 gray 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 [4] 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 [6].

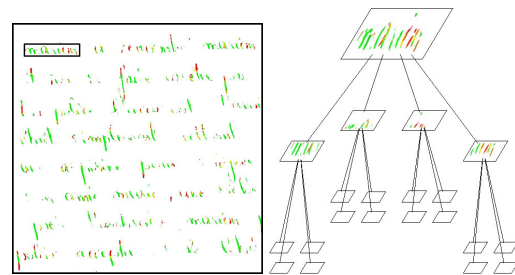
A mix of this kind of ideas can be found in [8]. The idea is to segment each documents into words with a technique given in [7], and then to resize and realign each word so that it can be easily compared with other images. Distance between images is then computed with the DTW distance on a set of features 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 [5] 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 [2] have tested appearance-based features for writer independent handwritten text recognition and compared it with heuristic features. Terasawa et al. in [9] 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 request in documents (see figure 2(a)). This could have been done by Gabor filters but our system has been designed in a genericity idea so that the Curvelet transform gives us a better analysis for the CBIR aspect ([3]) and, as we hope, for the clustering aspect.



(a) First (green) and second (red) orientations of the request detected in document (b) Quadtree decomposition of the two first orientations of the request

**Figure 2. Orientations detected in document and quadtree associated**

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 quadtree applied on the window in which each leaf contained the amount of pixels which have the searched orientation (see figure 2(b)). This means that, in each leaf, we keep how many pixels have the same predominant orientations as the request. Those numbers give information on where those predominant orientations are located in the window. We have tested a very naïve distance which consists in computing the ratio  $R = W/Q$  where  $W$  is the amount of pixels in one leaf for the sliding window and  $Q$  is the same amount but for the request. If  $R \geq 0.8$ , we consider the window as similar enough and we can get deeper in the quadtree. On figure 3 we have drawn the windows kept at three different stages. The lighter the window is, the deeper in the quadtree the window is considered.

Once the search is done in the original document, we do exactly same search in other documents of the database. Nothing is changed to the request characterization or to the new document to normalize those ones.

The next step is to test a more sophisticated distance

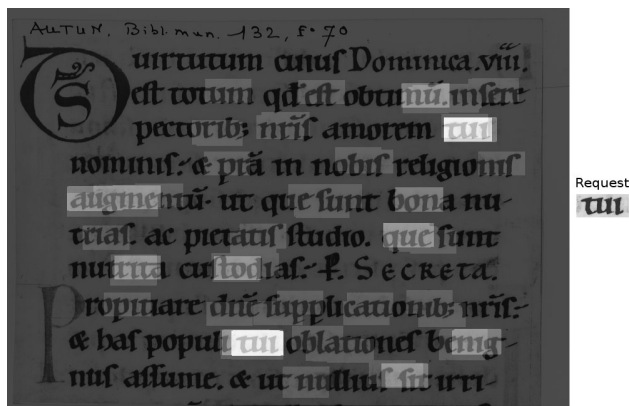


Figure 3. Quadtree distances representation over the original document

measure such as the DTW distance which seems to give good results in this kind of application. Until this test, very good results are shown by this naïve approach like it is shown in part 4. In most cases every occurrences of a word are retrieved with only three levels of decomposition in our quadtree.

#### 4. Discussion and conclusion

We have tested our approach on two kinds of documents: medieval and humanistic manuscripts. We have selected a small part of our databases in order to keep documents which have words statistically well represented. As we can see in table 1, results are really good for short words (such as Orders in George Washington’s manuscripts) but clearly decrease with long words (such as Companies). This is mainly due to the fact that our naïve distance measure is rigid and variability with long words is more frequent. It means that we have opposite results compared to [5] or [7] which are typically better when searched words are long. In the same idea, on contrary to previous work, we do not need to have a high regularity in handwritings but only a link on predominant orientations and their local densities.

	Recall	Precision	F-Measure 2PR/(P+R)
Humanistic (Short words)	0.79	0.71	0.75
Humanistic (Long words)	0.67	0.12	0.2
Medieval	0.85	0.67	0.75

Table 1. Recall/Precision results

The rigidity problem can be solved in almost two

ways. First way could be to use a less rigid distance measure between quadtrees such as, for example, a DTW distance. The second way could be to use a scaling window around region in the document which have a similarity with the query image in terms of organization of orientations. This second solution implies a step of localization and analysis of orientations in documents but this step could improve the execution time in the way that not all positions should be tested. For the moment, for a 900x900 image, it needs about 20 minutes on a Core2Duo T7100 to have results shown in part 3.

Without any optimization of our approach we obtained very good results compared to previous works which often need to segment (for example the words in [8]) and for the moment, all good responses have been found in the thirty first responses of our system. Compared to some very distant rank (as the rank 161 in the work mentioned in [5]), our responses ranks are very satisfying.

We have designed a generic tool which can be used at several levels of analysis (layout, CBIR, clustering and shape/word spotting) which has proven its performance in the CBIR field ([3]) and which give good results in a first step for the word-spotting. All these good results have been computed without any segmentation or binarization and there is no language dependency.

#### References

- [1] L. Boubchir and J. Fadili. Bayesian denoising based on the map estimation in wavelet-domain using bessel k form prior. In *ICIP05*, pages I: 113–116, 2005.
- [2] G. A. Fink and T. Plotz. On appearance-based feature extraction methods for writer-independent handwritten text recognition. In *ICDAR05*, pages 1070–1074, 2005.
- [3] G. Joutel, V. Eglin, S. Bres, and H. Emptoz. Curvelets based queries for CBIR Application in handwriting collections. In *ICDAR07*, pages 649–653, 2007.
- [4] A. Kolcz, J. Alspector, M. Augusteijn, R. Carlson, and G. Popescu. A line-oriented approach to word spotting in handwritten documents. *PAA*, 3(2):153–168, 2000.
- [5] Y. Leydier, F. Lebourgeois, and H. Emptoz. Text Search for Medieval Manuscript Images. *Pattern Recognition*, 12(40):3552–3567, Dec. 2007.
- [6] R. Manmatha and W. Croft. Word spotting: Indexing handwritten archives. In *Intelligent Multimedia Information Retrieval Collection.*, 1997.
- [7] R. Manmatha and N. Srimal. Scale space technique for word segmentation in handwritten manuscripts. *ScaleSpace99*, pages 22–33, 1999.
- [8] T. M. Rath and R. Manmatha. Features for word spotting in historical manuscripts. In *ICDAR03*, page 218, 2003.
- [9] K. Terasawa, T. Nagasaki, and T. Kawashima. Automatic keyword extraction from historical document images. *DAS06*, pages 413–424, 2006.