Evaluation of Handwriting Similarities Using Hermite Transform

Stéphane Bres, Véronique Eglin, Catherine Volpilhac-Auger

Evaluation of Handwriting Similarities Using Hermite Transform

BRES S.
LIRIS - UMR 5205,
INSA de Lyon,
Batiment Jules VERNE
69621 VILLEURBANNE CEDEX
stephane.bres@insa-lyon.fr

EGLIN V.
LIRIS - UMR 5205,
INSA de Lyon,
Batiment Jules VERNE
69621 VILLEURBANNE CEDEX
veronique.eglin@insa-lyon.fr

VOLPILHAC-AUGER C.
CERPHI – UMR 5037,
ENS LSH de Lyon,
15, parvis René Descartes - 69342 LYON CEDEX
c.volpilhac-auger@wanadoo.fr

Abstract

In this paper, we present a new method for handwriting documents denoising and indexing. This work is based on the Hermite Transform, which is a polynomial transform and a good model of the human visual system (HVS). We use this transformation to analyze handwritings using their visual aspect of texture. We apply this analysis to document indexing (finding documents coming from the same author) or document classification (grouping document containing handwritings that have similar visual aspect). It is often necessary to clean these documents before the analyze step. For that purpose, we use also the Hermite decomposition. The current results are very promising and show that it is possible to characterize handwritten drawings without any a priori graphemes segmentation.

Keywords: Document restoration, handwriting analysis, handwriting indexing.

1. Introduction

There are many different kinds of databases around the world, and all of them have to deal with the same problem, what ever information they hold: how to organize this information cleverly and how to retrieve visually similar information. It is a great challenge and it can not be resolved with a unique generic solution but it must be adapted to each kind of information. In this paper, we are working on handwriting documents corpus. Our purpose here is to characterize precisely handwritings whatever their authors are and to classify them into visual writers' families. Our approach considers handwritings as special drawings that create a specific texture we want to analyse by considering orientations at different scales. Orientations are considered as sufficiently relevant perceptual features to characterize the special texture of handwritten drawings. These orientations information are extracted by using the Hermite transform which is a particular polynomial transform and a good model of the receptive field profiles of the human visual system. This model leads to the development of an original method of handwriting classification by the computation of handwritings signature and similarity measures that reveal their "visual textural aspects".

1.1. Patrimonial handwritings documents

The databases we have to treat contain historical handwritings documents and the characteristics of these documents have a direct influence on the approach we choose for our orientations extraction. Many digital images of documents and more specifically, ancient manuscripts are degraded by the presence of strong artefacts in the background (see figure 1). This can either affect the readability of the text and, in our case, it compromises a relevant handwriting characterization. Consequently, most of the time, it is very difficult to directly extract the handwritings in those images. It becomes necessary to pre-process the images with a cleaning and denoising first step.

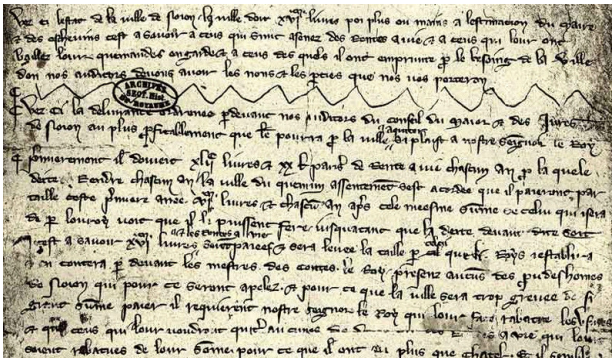


Figure 1. Examples of ancient manuscripts degraded by strong artefacts or spots [3].

Most generally, we consider the documents as a mixed signal composed by a textured background with a superimposed high frequency handwriting signal. Thresholding techniques are often not effective since the

intensities of background can often be close to those of the foreground text. Some approaches for text and background separation have been proposed in [5] where multistage thresholding techniques have been investigated to segment parts. Other techniques based on adaptative filtering have been tested on forensic documents to separate homogeneous textured background from handwriting marks, [2]. Some approaches consider a physical model of degradation to propose a mathematical model for text enhancement and background cleaning, [9]. In [12], the authors propose a decomposition of the signal into two blind sources where the overlapping texts and the supports (paper) texture are the unknown sources to be recovered with the consideration of different spectral bands of the documents.

1.2. Human visual system and Hermite transform

Texture features extraction is usually performed by linear transformation or image filtering, [1,7], followed by some energy measures or non-linear operator application (e.g. rectification). In this paper, we focus on the multi-channel filtering (MCF) approach. It is inspired by the MCF theory for processing visual information in the early stages of the human visual system, [1,7], where receptive field profiles (RFPs) of the visual cortex can be modelled as a set of independent channels. Each of these channels is tuned on a specific orientation and frequency. The use of these filters leads to the decomposition of an input image into multiple features images. Each of these images captures textural features occurring in a narrow band of spatial frequency and orientation. Among the MCF models having the above properties, Gabor filters have been widely used in texture feature extraction, [4], image indexing and retrieval, [13]. Another model corresponds to Hermite filters of the Hermite transform [6] that agrees with the Gaussian derivative model of the HVS. It has also been shown analytically that Hermite and Gabor filters are equivalent models of receptive field profiles (RFPs), [6],[8]. However, Hermite filters have some advantages over Gabor ones, like being an orthogonal basis leading to information decorrelation and perfect image reconstruction after decomposition. This is the main reason why we are interested in this transform. Moreover, a discrete representation of Hermite filters exits (the Krawtchouk polynomials) with the property of separability for an efficient implementation.

2. Hermite transform

In this paper we present a method for image document cleaning and indexing based on the Hermite transform. It is exploited here to decompose the initial signal into different parts depending on their frequencies characteristics (high or low). Most of the time, the noise or degradations that appear on ancient documents have low frequencies characteristics, while the writing by itself is composed of high frequencies. It is of a great interest to separate them. This is exactly what we want to achieve with the Hermite transform. In the following paragraph, we present the definition of the Hermite transform.

2.1. Cartesian Hermite Filters and Krawtchouk Filters

Polynomial transforms are the decomposition of a signal $l(x,y)$ into a linear combination of polynomials. The original signal is locally treated, window by window. These windows are positioned on the signal with a constant translation step, and the polynomials are orthogonal with respect to this specific window shape. In the case of Hermite transform, the window $v(x,y)$ is a Gaussian window. Hermite filters $d_{n-m,m}(x,y)$ decompose the original signal $l(x,y)$ by computing a localized signal $l_v(x-p,y-q) = v^2(x-p,y-q)\ l(x,y)$ where $v(x,y)$ is a Gaussian window with spread σ and unit energy, into a set of Hermite orthogonal polynomials $H_{n-m,m}(x/\sigma\ ,\ y/\sigma)$. Coefficients $l_{n-m,m}(p,q)$ at lattice positions $(p,q)\in P$ are then derived from the signal $l(x,y)$ by convolving with the Hermite filters. These filters are equal to Gaussian derivatives where $n–m$ and m are respectively the derivative orders in x- and y-directions, for n=0,…,D and m=0,…,n. Thus, the two parameters of Hermite filters are the maximum derivative order D (or polynomial degree) and the scale σ . Hermite filters are separable both in spatial and polar coordinates, so they can be implemented very efficiently. Thus, $d_{n-m,m}(x,y) = d_{n-m}(x)\ d_m(y)$, where each 1-D filter is:

$$d_n(x) = \left((-1)^n \Big/ (\sqrt{2^n \cdot n!}\sqrt{\pi}\sigma)\right) H_n(x/\sigma) e^{-x^2/\sigma^2} \qquad (1)$$

where Hermite polynomials $H_n(x)$ are orthogonal with respect to the weighting function $exp(-x^2)$, and are defined by Rodrigues' formula in [6] by:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \qquad (2)$$

In the frequency domain, these filters are Gaussian-like band-pass filters with extreme value for $(\omega\sigma)^2 = 2n$, [8], and hence filters of increasing order analyze successively higher frequencies in the signal.

Krawtchouk filters [11] are the discrete equivalent of Hermite filters. They are equal to Krawtchouk polynomials multiplied by a binomial window $v^2(x) = C_N^x / 2^N$, which is the discrete counterpart of a Gaussian window. These polynomials are orthonormal with respect to this window and they are defined by :

$$K_n(x) = \frac{1}{\sqrt{C_N^n}} \sum_{\tau=0}^{n} (-1)^{n-\tau} C_{N-x}^{n-\tau} C_x^{\tau} \qquad (3)$$

for x=0,…,N and n=0,…,D with $D \le N$. It can be shown that the Krawtchouk filters of length N approximates the Hermite filters of spread $\sigma = \sqrt{N/2}$. In order to achieve fast computations, we present a normalized recurrence relation to compute these filters, see [8] :

$$K_{n+1}(x) = \frac{1}{\sqrt{(N-n)(n+1)}}\Big[(2x-N)K_n(x) - \sqrt{n(N-n+1)}K_{n-1}(x)\Big] \ , \ n \ge 1 \qquad (4)$$

with initial conditions $K_0(x) = 1$, $K_1(x) = \frac{2}{\sqrt{N}}\left(x - \frac{N}{2}\right)$.

2.2. Steered Hermite filters and Gabor-like Hermite Filters

In order to have a multi-channel filtering (MCF) approach based on Hermite filters, they must be adapted to orientation selectivity and multi-scale selection. For that purpose, we apply their property of steerability, [6,8]. The resulting filters may be interpreted as directional derivatives of a Gaussian (i.e. the low-pass kernel). Since all Hermite filters are polynomials times a radially symmetric window function (i.e. a Gaussian), it can be proved that the n+1 Hermite filters of order n form a steerable basis for every individual filter of order n. More specifically, rotated versions of a filter of order n can be constructed by taking linear combinations of the filter of order n.

The Fourier transform of Hermite filters $d_{n-m,m}(x,y)$ can be expressed in polar coordinates $\omega_x = \omega \cos\theta$ and $\omega_y = \omega \sin\theta$ as $\hat{d}_{n-m,m}(\omega_x, \omega_y) = \hat{d}_n(\omega)\alpha_{n-m,m}(\theta)$ where $\hat{d}_n(\omega)$, which expresses radial frequency selectivity, is the 1-D Fourier transform of the nth Gaussian derivative in (1) but with radial coordinate r instead of x. The cartesian angular functions of order n for m=0,…,n, are given as

$$\alpha_{n-m,m}(\theta) = \sqrt{C_n^m}\cos^{n-m}\theta \cdot \sin^m\theta \qquad (5)$$

which express the directional selectivity of the filter.
Steered coefficients $l_n(\theta)$ resulting of filtering the signal $l(x,y)$ with these steered filters can be directly obtained by steering the cartesian Hermite coefficients $l_{n-m,m}$ as:

$$l_n(\theta) = \sum_{m=0}^{n} l_{n-m,m} \cdot \alpha_{n-m,m}(\theta) \qquad (6)$$

scale representation that fulfils the desired constraints in the frequency domain, which are mainly the number of scales S (radial frequencies ω_0) and the number of orientations R in the filter bank. Since previous works have been done essentially with Gabor filters, we have then adopted a similar multi-channel design. Moreover, both Hermite and Gabor filters are similar models of the RFPs of the HVS [8]. For these reasons, we have named the resulting filters as Gabor-like Hermite filters. In summary, construction of a Gabor-like Hermite filter bank requires the following procedure. First of all, set the number of desired scales S and orientations R and for each of the scales s=0,…,S–1 compute :

- the radial central frequency ω_0 and the spatial spread σ_x of respective filters.
- Krawtchouk parameters such as window length N and filter order D.
- Krawtchouk filters: get the corresponding Krawtchouk polynomials through (4) and multiply them by a binomial window of length N.
- Input image convolutions with Krawtchouk filters to obtain cartesian coefficients.
- Steering coefficients to desired orientations through (6) and (5) to obtain the equivalent multi-channel outputs.

3. Patrimonial documents denoising

Our proposition uses the Cartesian Hermite transform (computed throw the Krawtchouk filters [11]) that extract the local frequencies of a signal. Figure 2 presents the Hermite decomposition of a document at a given scale N=16 and up to degree 2. The N value defines the size of the array of interest, and thus, the size of the filters we use for denoising. It is correlted to the resolution of the treated document. The most top left image is equivalent to a Gaussian low pass filtered image. Using the higher degrees in both directions allows extracting high frequencies of the original image. As we explain earlier, low frequencies contain information on the background and high frequencies contain information on the writings we want to keep, if their levels are sufficient (above a certain threshold).

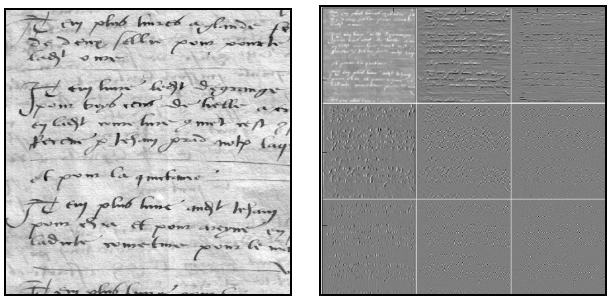


Figure 2. 2D - Hermite transform using Krawtchouk filters for N=16 and up to degree n=2 for the rows and the columns

The first step of our denoising process is then to decompose the original image using the Hermite decomposition. In the second step, we reconstruct the

high pass image I_H using degrees higher than N/4 in both directions. Too small values are filtered at this step (values less than 10% of the maximum). We obtain an image with a cleaner background: low frequencies and small high frequencies variations have disappeared (see Figure 3.1). A more detailed view is presented on figure 3.4. This image I_H is used as a mask that localizes the writings we want to keep. Pixels belonging to the background have low values in a low values neighbourhood. Pixels belonging to writings have a highly contrasted neighbourhood. An example of document denoising is shown on figure 3.2. The original image is presented figure 2. Details of these images are presented on figure 3.3 and 3.5. Using denoised images allows focusing on the handwritings by itself.

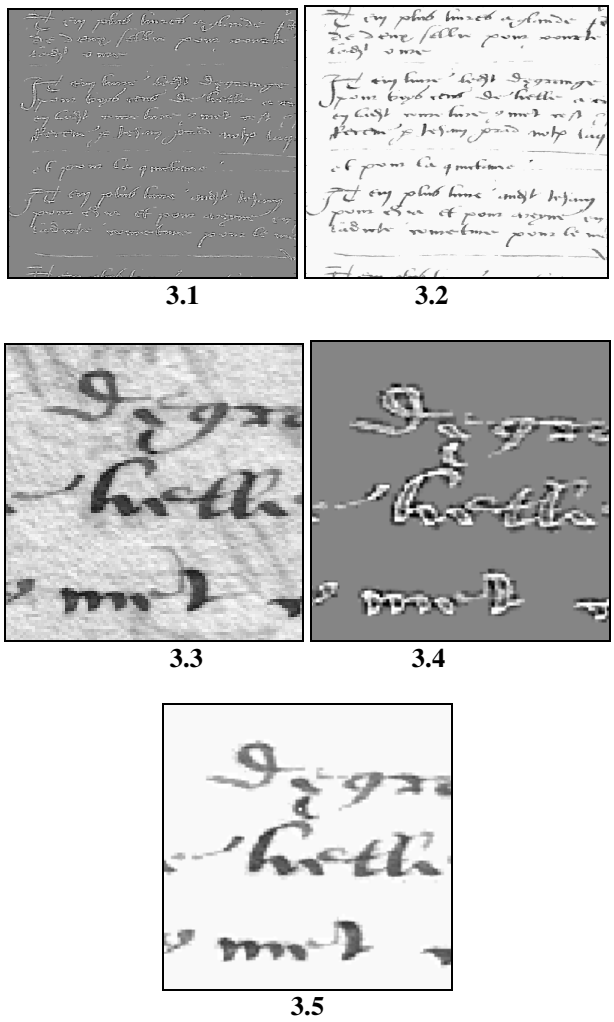


Figure 3. Example of document denoising. High pass image based on Hermite decomposition (3.1) - Denoised document (3.2) - Detail of original document (3.3) - Detail of high pass image (3.4) - Detail of denoised image (3.5).

The main difference between this Hermite based approach and a classical adaptative thresholding comes from the local frequency decomposition we make here. Most of the time, adaptative thresholding methods classify pixels as background pixels or foreground pixels (handwriting lines) depending on local statistical values. Our method uses local frequency decomposition. On their principle, these methods are different, because Hermite based approach allows filtering that take into account the local frequency to decide between as background pixels or foreground pixels. This could be especially interesting in case of wide degradation areas containing black low frequencies. Such areas will not respond to high frequencies Hermite filters but adaptative thresholding can still keep some pixels as foreground pixels.

4. Handwritings document signature

Handwriting characterization will be done using orientations extraction by Gabor-like Hermite filters. The two parameters (number of scales S and the number of orientations R) needed for Gabor-like Hermite filters are fixed to S=4 and R=6. This will lead to 24 oriented filters. For a given pixel, each of these 24 filters will give responses that characterize a given orientation at a given scale. We only keep responses on pixels identified as handwritings lines pixels in the denoising step. Then, we have a 24 values vector for each handwritings lines pixel. All these vectors can be represented as a cloud, in a 24 dimensions space, which is a good characterization of the analyzed handwriting. Unfortunately, these signatures are far too big to be used directly, and have to be reduce to something as small as possible with a minimal information loss. We choose to keep geometrical information of the clouds, like their gravity center (mean values on each of the 24 coordinates) and main axis (eigenvectors and eigenvalues) after an PCA-like step. Our signature for a given handwriting document is then the 24 means values of the results coming out of the filters bank, 24 normalized eigenvectors and the 24 corresponding eigenvalues of the covariance matrix computed from the centered cloud of orientations vectors. Moreover, experiences show that we do not need to keep all the eigenvectors and associated eigenvalues : only the 3 or 4 greater values need to be stored.

5. Handwritings document indexing

We need to define a distance between these signatures to introduce the similarity notion in the database. Similarity leads to indexing which is the goal we want to reach. With a similarity measure, it is easy to build an indexing motor that can classify the documents and retrieve the most similar documents to a requested one.

5.1. Similarity computation

In practice, our signature for image number *i* is made of 24 mean values M_i (n), 4 eigenvalues and the 4 normalized eigenvectors V_i corresponding to the 4 greater eigenvalues L_i. L_{io} quantifies the importance of

the vector V_{io} in the shape of the cloud. The distance D we choose to define uses both information of mean values M_i and the couples vectors V_i and values L_i. This distance D is the combination of the distance D_M between the mean values M_i and a multiplicative normalized coefficient $\overline{D_E}$ coming from the eigenvectors and eigenvalues. The D_M (Hi, Hj) distance between handwriting *i* and handwriting *j* is defined by :

$$D_M(H_i, H_j) = \sum_{n=1}^{24} \left| M_i(n) - M_j(n) \right| \qquad (7)$$

The multiplicative normalized coefficient $\overline{D_E}$ coming from the eigenvectors and eigenvalues is based on the non normalized distance D_E between weighted eigenvectors. The weights we use here are their corresponding eigenvalues :

$$D_E(H_i, H_j) = \sum_{n=1}^{4} \left| L_i(n).V_i(n) - L_j(n).V_j(n) \right| \qquad (8)$$

We obtain $\overline{D_E}$ after a normalization step. For that purpose, we divide D_E by its maximum value to have a value between 0 and 1. Thus :

$$\overline{D_E}(H_i, H_j) = D_E(H_i, H_j) \;\; / \;\; \sum_{n=1}^{4} \sqrt{L_i^2(n) + L_j^2(n)} \qquad (9)$$

Finally, the distance $D(H_i, H_j)$ between handwriting H_i and handwriting H_j is can be expressed as :

$$D(H_i, H_j) = D_M(H_i, H_j) \,.\, \overline{D_E}(H_i, H_j) \qquad (10)$$

This distance is symmetrical, which is a good property to assure coherent results during multiple comparisons of databases documents. A small distance means high similarity.

5.2. Practical results

We have tested the whole system on our personal database composed of documents coming from different authors but mainly patrimonial handwritings documents. Most of the time, we have full pages of the same author and for evaluation purpose, these pages are divided into smaller images, 9 per page. Then, most pages give us 9 images from the same author, containing what we can suppose to be similar handwritings. This is how we build our "ground truth": images coming from the same original page image should look the same and have similar handwritings. It is difficult to complete this ground truth with similarities between different author's handwritings because of the subjective judgment involved in such estimation. Figure 4 gives some examples of images coming from the same original page. Our database contains 1438 images coming from 189 different authors, in different languages and alphabets.

To illustrate the discrimination possibilities of our signatures, we present, on figure 5, a 2D representation of the 24D signature space. This 2D representation is obtained by PCA on the mean values M_i of the text images of figure 4. The 9 points in the ellipse #n correspond to the 9 text images of the author presented in the line #n of figure 4 (only 3 of these images are shown here).

Figure 4. Examples of images coming from the same authors (one author per line)

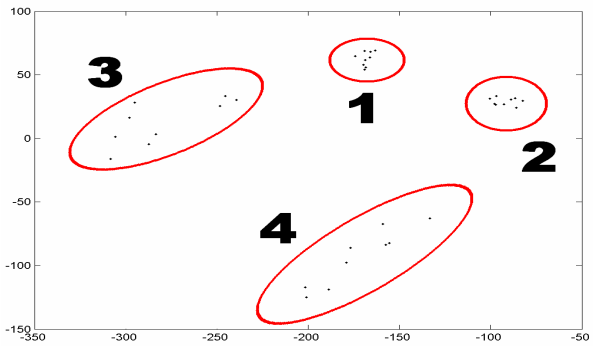
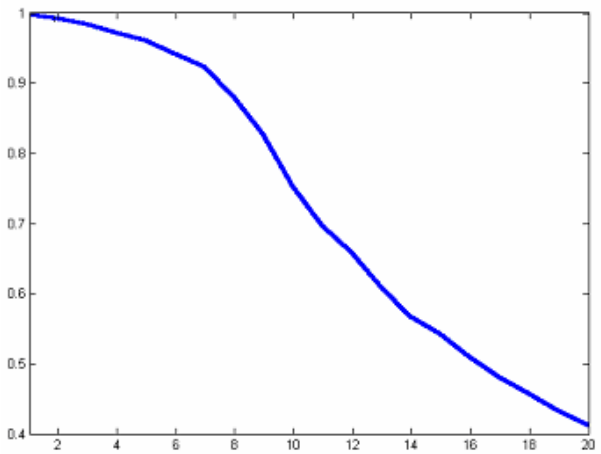


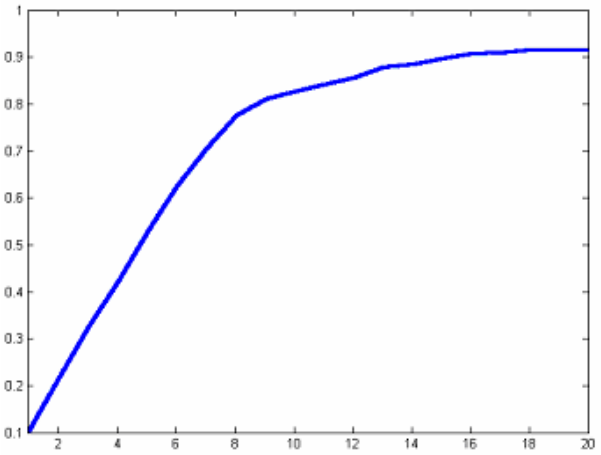
Figure 5. 2D projection obtained by PCA of the 24D signature space. Each point is the mean value of a text image. They are grouped by author presented on figure 4: ellipse #n for the author of line #n

The global results we obtain are really promising because, according to our ground truth, a given request has in the ten first better answers (documents with the higher similarity or equivalently the smaller distance)

in average more than 83% of correct responses, see recall curve on figure 6. This is an average value computed on the documents that have 9 similar images in the database. These precision and recall curves are a common way to show the efficiency of an indexing system. They have been computed using the 20 first responses. Let's remember that we only have 9 images for each handwriting. That is the reason why the precision decreases strongly after the 9th response.



6.1



6.2

Figure 6. Classical Precision curve (6.1) and Recall curve (6.2) computed on the entire database containing more than 1400 handwriting documents

6. Conclusion

This work is a response to scientific problems of historical handwritten corpus digitalization. It deals with the handwriting denoising and indexation and is applied here to a multi-language and multi-alphabet corpus. We propose here a biological inspired approach for images denoising (by a background *cleaning*) and handwriting characterization for corpus indexing. The developed perception based model lies on the Hermite frequencial decomposition for image denoising and indexing. Our motivation is directly linked to the difficulty to perform efficient image processing on degraded handwriting historical documents without a priori knowledge on the image content. In that way, we have chosen a segmentation free approach that is global and generic. The current results of handwriting denoising and classification with orientation Hermite based features are very promising. We are currently working on an enlarged database in connection with recent digitalization European project.

7. References

[1] Bovik, A.C., Clark, M., and Geisler, W.S., Multichannel texture analysis using localized spatial filters, *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 12, pp. 55-73, 1990.

[2] Franke, K., Koppen, M., A computer-based system to support forensic studies on handwritten documents, Int. Jour. Doc. Anal. Reco., 3:218-231, 2001.

[3] GALLICA : digital library of the BNF : http://gallica.bnf.fr

[4] Grigorescu, S.E., Petkov, N., and Kruizinga, P., Comparison of texture features based on Gabor filters, *IEEE Trans. Image Processing*, vol. 11, pp. 1160-1167, 2002.

[5] Leedham G., Varmas,S., Patnkar, A., Govindaraju,V., Separating text and background in degraded document images- a comparison of global thresholding techniques for multi-stage thresholding. In Proceedings of the 8th inter. Workshop on frontiers in handwriting recognition, 2002, pp. 244-249.

[6] Martens J.-B., The Hermite transform – Theory, IEEE Trans. Acoust., Speech, Signal Processing, vol. 38, no. 9, pp. 1595-1606, 1990.

[7] Randen, T. and Husøy, J.H., Filtering for texture classification: A comparative study, *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 21, pp. 291-310, 1999.

[8] Rivero-Moreno C.J., Bres S., Conditions of similarity between Hermite and Gabor filters as models of the human visual system, In Petkov, N.& Westenberg, M.A. (eds.): Computer Anal. of Images & Patterns, Lectures Notes Computer Sci., vol. 2756, pp.762-769, 2003.

[9] Sharma, S. Show-through cancellation in scans of duplex printed documents. IEEE Trans. Image Process, 10(5):736-754.

[10] Strouthopoulos, C. and N. Papamarkos, *Image and Vision Computing,* Text identification for document image analysis using a neural network,pp879-896, vol.16, 1998.

[11] Swarttouw R.F., Koekoek R., *The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue*. Delft University of Technology, Faculty of Information Technology and Systems, Department of Technical Mathematics and Informatics, Report no. 98-17, The Netherlands, 1998

[12] Tonazzini A., Bedini L., Salerno E., Independant component analysis for document restoration, Inter. Jour. on Doc. Anal. and Reco., vol.7, N°1, 2004, pp. 17-27.

[13] Wu, P., Manjunath, B.S., Newsam, S., and Shin, H.D., A texture descriptor for browsing and similarity retrieval. *Signal Proc.*: *Image Comm.*, vol. 16, no. 1,2, pp. 33-43, 2000.