Compression of Handwriting Images : a way to define a writing style

Nicole VINCENT and Stéphane BARBEZIEUX

LI / E3i - Université de TOURS
64, avenue J. Portalis
37200 TOURS

vincent@ univ-tours.fr
tel : 33 2 47 36 14 12 fax : 33 2 47 36 14 22

Abstract

The final aim of the study is to obtain a clustering of off line writings that would improve the efficiency of the recognition methods. Here, we propose the definition of new parameters that give some idea on the global aspect of a writing.

An image fractal compression is relying on inner autosimilarities. The quantification of the quality of a compression involves several parameters, the most important is compression rate. Besides, comparison between an image and what is obtained after a compression / decompression process can be achieved through comparison of the pixel grey levels. The quadratic error mean is computed. The value of this parameter is low when the image has good autosimilarity properties.

A study of the behaviour of these parameters has been performed using many different samples of writings. In particular the height of the writing, the spacing and the shape of the letters have been made vary. So it has been possible to understand what characteristics are influencing these parameters. The influence of the ink color has been studied too. The hypotheses have been verified on very regular, simulated images as well as on printed text, where the modified variables are more easily controlled.

Keywords

handwriting analysis, global analysis, text compression, fractal compression.

Résumé

En vue d'établir un classement de l'écriture qui permette une amélioration de l'efficacité des méthodes de reconnaissance, nous proposons la définition de nouveaux paramètres qui donnent une idée de la régularité globale d'une écriture.

La compression fractale d'une image repose sur la présence d'autosimilarités internes à l'image. La mesure de la qualité d'une compression donne lieu à divers paramètres, en particulier le taux de compression. Par ailleurs, la comparaison entre une image et l'image obtenue après compression puis décompression donne lieu à un calcul de la moyenne des erreurs quadratiques induites sur les niveaux de gris des pixels. La valeur de ce paramètre est d'autant plus faible que l'image possède une forte propriété de similarité interne.

Une étude du comportement de ces paramètres sur des échantillons variés d'écriture, où l'on a plus particulièrement envisagé des variations de la taille des espaces, de la forme des lettres, permet de mettre en évidence les caractéristiques discriminées par ce type de paramètres. Une étude de l'influence de la couleur de l'écrit a été menée. Les hypothèses faites ont été vérifiées sur des images régulières de simulations d'écriture ainsi que sur du texte imprimé pour lequel les valeurs des variables sont mieux maîtrisées.

Mots clés

analyse de l'écriture manuscrite, analyse globale, compression de texte.

Introduction

As a matter of fact, the best way to compress the image of a text is to achieve its understanding and to replace the image by an ASCII text. Nevertheless, yet no universal technic has been proposed to fulfill this aim. The problem of text recognition and more specifically the problem of handwriting recognition is a complex one. One of the difficulties comes from the great variability encountered among the different styles of writing. The patterns involved in a forward sloping writing are not the same as those present in a printed handwriting. In this paper, we are dealing with the problem of finding parameters that allow a first coarse clustering of handwritings. We propose to take advantage of an image compression method and more precisely of a fractal compression method.

The variability of writings has been studied by several teams [CRE95 , LER96], the analysis of the writing has been too[PLA93 , PLA97]. In fact, each piece of writing depends upon the personality of the writer, on his mood too, on the moment of the day he is writing, and also on physical conditions, the type of pen that is used or the dimension of the space the writer has at his disposal to write.

The knowledge of a global family which the writing would belong to, would enable to adapt the recognition method to the patterns involved. The recognition technic as well as the learning set could be chosen according to the writing style and then the recognition rate might be improved.

Our aim is to extract some information about the writing image more than about the writer himself. So we have chosen not to use a binarised image of the text, but rather a grey level image, with a 200 dpi resolution. The difference between the image and the original text is the most obvious in the case of a text written with a pencil. Very often some holes appear on the image within the lines, but human eye does not see them. Human vision is far more efficient than artificial vision when dealing with interpretation of visible scenes.

Besides, any human being has no difficulty in classifying texts according to the writing style, and this is done without any effective reading of the text. The evidence is induced by a global perception of the text, rather than by a tidy and conscious study of the written lines. So we have chosen to consider the inner similarities that can be detected in the image of writing. Doing so, we go apart from all the methods that can be said descriptive [CRE95 , LER96] and that are looking either for some specific patterns or for some characteristics such as a local orientation. Our approach is a global one.

A writing is the succession of strokes that are drawn from the action of the writer muscles. The curvature radii are for instance depending, on the personality, on the ability of the writer and on the writing speed. Some studies have been achieved on this subject[PLA93 , PLA97]. Some invariants can be defined, that are quite convenient for writings. In fact, some similar patterns can be seen, at different scales, in various letters or embellishments or even ligatures. Their orientations may differ, but the patterns remain similar. So we think the study of the similarities present within the image, will help us to describe a global perception of the image. Nevertheless, no specific pattern such as loops has to be looked for. Globally, it can be thought that if many similarities have been found, the regularity of the writing would be greater, otherwise the writing might be more erratic or more degenerated.

A recent study has shown that a global feeling can be extracted from a writing using one or two parameters. More precisely, the fractal behaviour of writing has been justified [BOU97]. This has led us to look for autosimilarities in the image and to consider a text as an image, applying to it a fractal compression followed by a decompression step. This allows to define some new parameters. These will give some new knowledge that permits a better discrimination of the different styles. This new approach uses all the information contained in the image, in particular the grey levels of the pixels. A binarisation step is not needed as was the case when computing the writing fractal dimension [VIN95].

Our paper is organised into three sections. In the first part, we recall the principles of fractal compression, and then we present the parameters we have defined from the compression of the handwriting image and we have taken into account. Finally, in the last part, we comment some results obtained from the use of these parameters.

1. Fractal compression principles

At first, the use of fractals has been developing in the field of synthesis images. Some simple transformations of the plane were allowing, after several iterations, to obtain realistic images with much details as, for example, the ferns. There, the main property involved is autosimilarity. The transformations were in fact the IFS (Iterated Function Systems) [BAR88]. In fact, each transformation is an affine mapping and can be defined by the 6 coefficients that appear in the matricial formulation. Then comes the idea of solving the inverse problem : starting from an image, is it possible to find the associated transformations, those that enable to reconstruct the image by an iterative method ?

To achieve fractal compression [BAR93 , FIS95], an initial image I_0, that is known by its pixel grey level list, is replaced by a transformation defined on the set of all possible images and this mapping has the

A splitting into triangles is more complex than the previous method, it allows the definition of more homogeneous zones that better respect the contours present on the image. More generally a splitting into quadrilaterals may be considered by the merging of adjacent triangles that have a high similarity. This allows to reduce the number of the transformations that are to code the image, and so the compression rate is increased.

In case of text images, the contrast between text and background is most often well defined. So we thought to make some similarity research easier by extracting some background zones, that is to say more uniformed zones.

Writing presents approximately two privileged directions, horizontal and vertical, so we have opted for a splitting into rectangular zones as much as possible centered on the text.

Hence, the splitting of a zone into two sub-domains is made here according to a vertical or an horizontal straight line that permits to minimize the variance of the grey levels in each zone. So, domains having no written text rapidly appear, they are assumed to be uniform zones. For the ones that do not involve written text, it will be easy to find an homothetic spatial transformation and a grey level translation to code the domains. In the zones containing text, the subdivision will be smaller, hence in these zones transformations will be more precise.

A more sophisticated research of domains more adapted to the writing would have led to a local analysis of the writing, a thing we wanted to avoid. So we have opted to that compromise, using only rectangular zones mostly concerned with either text or background.

2.2. w_i choice

Two parts are concerned, on the one hand the definition domain, on the other hand the mapping explicit shape. It has already been specified we are restricting the possible D_i to domains similar to R_i with a ratio equal to 2. As possible D_i, we may choose among domains covering one an other or consider only those that would constitute a division of D. Then computation time is highly penalized by an exhaustive search among all the possible D_i.

Here we only consider 8 spatial transformations (four rotations including identity and four compositions of symmetry and rotation); for each possible transformation, s and o may be calculated through minimization of the quadratic error done by replacing the grey level of the pixels in R_i by the values obtained applying the w_i transformation to the D_i domain. True grey levels are replaced by transformed grey levels. When the "best" transformation has been calculated, the error, the minimum error, permits to estimate the transformation quality. It is this value that will lead to accept R_i as a subdivision or that will oblige to realize a more precise splitting, so that error is reduced. Of course, a R_i minimum size has been previously fixed.

Here no specific choice is induced by the fact we are working on text images.

2.3. Compression rate

The first parameter we shall use is concerning writing ability to be highly compressed by a fractal method. It depends on the presence of many similarities according to the 8 considered mappings.

The determination of the w_i, allows to code the image by this list of transformations that defines w, and the initial image is the fixed point of the global transformation.

The compression rate is the ratio of the number of bits necessary to code the grey levels of the image pixels to the number of bits needed to code the transformations and also the image division. As the splitting is processed in a recursive way, it is enough to note the way it has been obtained. As the size of the studied images is limited to 2048x2048 pixels, and N_t being the number of sub-domains R_i, the compression ratio is equal to

$$\frac{8\,(b-a+1)\,(d-c+1)}{49\,N_t-13}$$

On writing images, it is frequent that some parts of the scanned image do not contain any text. Now, we are in fact interested by properties of the writing itself and not by the writing global disposition on the page. Our method and results should be valid whatever the studied text may be. Therefore, we have decided to define text compression ratio rather than image compression ratio. This last parameter will be chosen later on. Here, for each mapping we shall note the D_i and R_i positions. Then, noting N_e the number of R_i containing writing, we define compression ratio of writing noted T_e by :

$$T_e = \frac{8\,(b-a+1)\,(d-c+1)}{78\,N_e}$$

I_0 image as a single fixed point. The theory is based on the fixed point theorem valid for contractant mappings in a complete metric space.

A gray level image, I_0, can be seen as an element of E=[a,b]x[c,d]x[0,N-1], where N is the number of possible gray levels. The first two coordinates concern the spatial position of the pixels in the image, D=[a,b]x[c,d] and the third one concerns the gray level of the pixels. Then, the initial image can be denoted $DxI_0(D)$. The transform that is to be found, is defined from E onto itself. The calculus of an approximation of the ideal mapping leads to a non reversible fractal compression technic, that is to say some information will be lost when the decompression process is achieved.

In part, the approximation is due to the limitation we have to put on the considered local transformations. Besides, to improve the quality of the compression process, the global transformation, w, is looked for in a local way, $w = \cup w_i$. More precisely, the image can be seen as the union of separated zones $R_ixI_0(R_i)$. For each zone, two elements are to be found. A counter domain that has the form $D_ixI_0(D_i)$ is looked for and also, the mapping itself w_i that will enable the best reconstruction of the initial image. Of course, D is equal to $\cup R_i$ but is not necessarily equal to the union of the D_i.

The w_i transform has two parts, one is relative to the spatial aspect, and maps D_i onto R_i, the other concerns the grey levels and maps $I_0(D_i)$ onto $I_0(R_i)$.

In order to simplify the search for the best transform, among possible geometrical transformations are only rotations, symmetries, translations and homotheties for which a matricial representation is easily considered. Besides, the mapping of the grey levels will be taken into account. Then the matricial formulation of the map can be formulated by :

$$\begin{pmatrix} x' \\ y' \\ n' \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s \end{pmatrix} \begin{pmatrix} x \\ y \\ n \end{pmatrix} + \begin{pmatrix} e \\ f \\ o \end{pmatrix}$$

From this form, it can be noticed, that s allows to control the contrast whereas o deals with the luminosity.

In order to define the part of the transform that is necessary for us

$w : E \to E$ we define $w : \cup D_i \times [0,N-1] \to E$

The first step is to choose the division of D in subsets R_i. Then to each R_i, a D_i and a local transformation w_i that is computed from the original image have to be associated. Very often, in order to be sure the mapping is a contractive one, , the domain D_i is chosen similar to R_i with an homothetic ratio equal to 1/2. Among all the possible D_i, the " best " one will be associated to the " best transform ", that is mapping in the most exact way $D_ixI_0(D_i)$ onto $R_ixI_0(R_i)$. The distance between $w_i(D_ixI_0(D_i))$ and $R_ixI_0(R_i)$ has to be minimized.

By this process, the I_0-mapping is approximated in such a way that I_0 is the fixed point of the transform w. This fixed point can be obtained as the limit of a sequence of images. Starting from any initial image, w(I) is built and then, the successive iterations. The quality of the transform depends upon the choice of the R_i, upon the choice of the D_i, upon the set of the local mappings that are to be considered, upon the distance that has been chosen to quantify the quality of the rebuilt image, it also depends on the properties of the initial image itself, and on the quality of the inner similarities that are contained in the image.

As the result of the compression process, must be kept in memory, the location of the D_i, the location of the corresponding R_i as well as the elementary transforms.

After the most important lines of the fractal compression method have been given, we may see many choices are still to be done, they are linked to each type of image. Next paragraph will deal with the adaptation of the method to the case of handwritten text images; then it will be shown how it is possible to define parameters that give some idea on the characteristic properties of the handwriting images.

2. Adaptation to handwritten texts

2.1. R_i choice

The most simple solution to choose the subdivision of D into the R_i is the subdivision using squares all of the same size, for example 4x4 squares.

Many improvements have been brought to these types of partition. For instance splitting by quadtree methods that splits a domain into four sub-domains of the same size if the " best " local transform that has been found from the large domain is not good enough. The splitting can be done, at each step, either through a vertical or an horizontal scanning. This gives some rectangular domains and a better adaptation to the image.

2.4. Error quadratic mean

This first parameter associated with compression ratio, measures writing capacity to be more or less highly compressed. The second is intended to measure the quality of the obtained compression. Decompression step consists in the determination of the fixed point of the transformation that has been stored. The classical demonstration of fixed point theorem gives an easy way to find it. Any point in E can be chosen and successive images are computed. They are obtained applying several times the w application defined as it is in 2.1. and 2.2. The fixed point is the limit of this image sequence. In practice, we stop computation when two consecutive images are close enough and we have noticed, in our case about ten iterations give good results.

The comparison of the initial image I_0 and of the decompressed image I_1, allows the measurement of the compression quality. The two images have the same size. Let us note Np the number of pixels, n_{0i} and n_{1i} the pixel grey levels respectively in the two images. Then, the quadratic error mean is defined by :

$$\text{eqm} = \frac{1}{Np} \sum_{i=1}^{Np} (n_{0i} - n_{1i})^2$$

If the scanned image uses 256 grey levels, the values of this parameter are varying in the interval [0,255].

For an equivalent purpose, and very often used, the ratio signal to noise may be considered. Too it gives an idea of the retrieved image quality. The higher it is, the higher the quality of the retrieved image is. Actually, a value about 25 to 30 is correct. This ratio, noted " psnr ", is expressed in decibels.

$$\text{psnr} = 10 \log_{10}\left(\frac{255^2}{\text{eqm}}\right)$$

It may be computed, either on the whole image, or only on the zones containing some pieces of writing. It is this last context that appears to be more significant for our study, as it has already been specified concerning compression ratio.

These two parameters being defined, we shall now study their behavior according to the different characteristics a handwriting may present.

3. Application to writing

The results that are presented here, are concerning the study of text images that have all been digitized with a 200 dpi resolution. Of course, in order to obtain results independent from the semantic meaning of the written text, we have experimentally observed, that the stability of the results, that are in fact relying on a statistical study, would be ensured only if the length of the processed text is long enough. Actually it must not be limited to one or two words, but rather take into account 6 to 10 centimeter long line. From an intuitive point of view it can be thought that for a large image, the search of similarities is done on more examples, hence eqm would have a natural tendency to decrease. So, its stability shows in fact a writing statistical homogeneity.

Precise studies have been led, dealing with the effect of resolution, of the space width that may occur between letters, words or lines, with the effect of writing slant and of line orientation, of writing conditions, in particular of ink color. In the following paragraphs, we give a rapid overview of some results that have been obtained.

3.1. Writing representation

Before we present some more precise results, we may observe the dispersion of the values we have obtained while studying a set of about 40 writings. The dispersion will make the discrimination power of the parameters apparent. Compression rate is between 1.8 and 2.4. As for psnr, it is varying between 5 and 50. Graph on figure 1 gives text compression rate vs. psnr. It may be noticed that small narrow writings gather on the left of the graph whereas quite rounded writings, where loops are well formed, give rise to a higher compression rate.

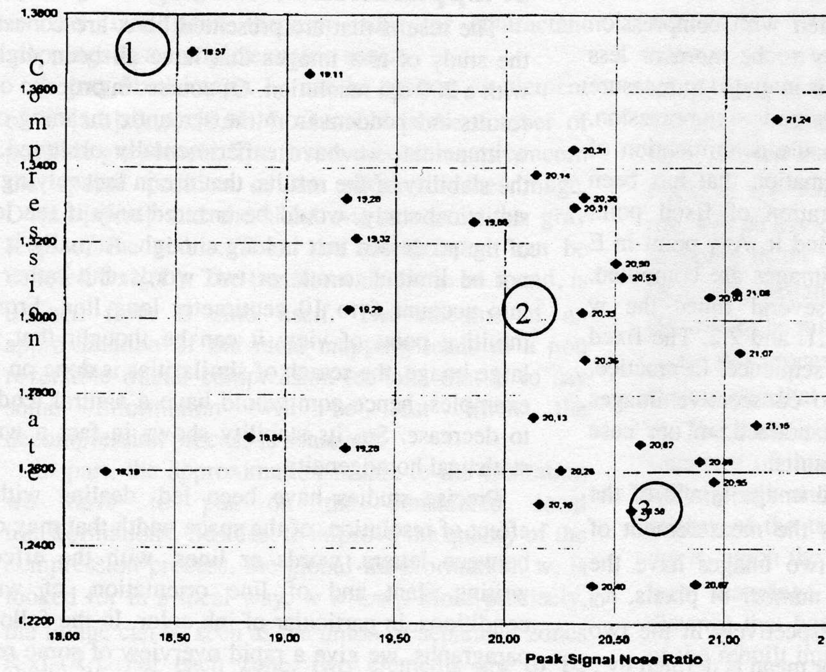

Figure 1 : Compression rate vs. Psnr.

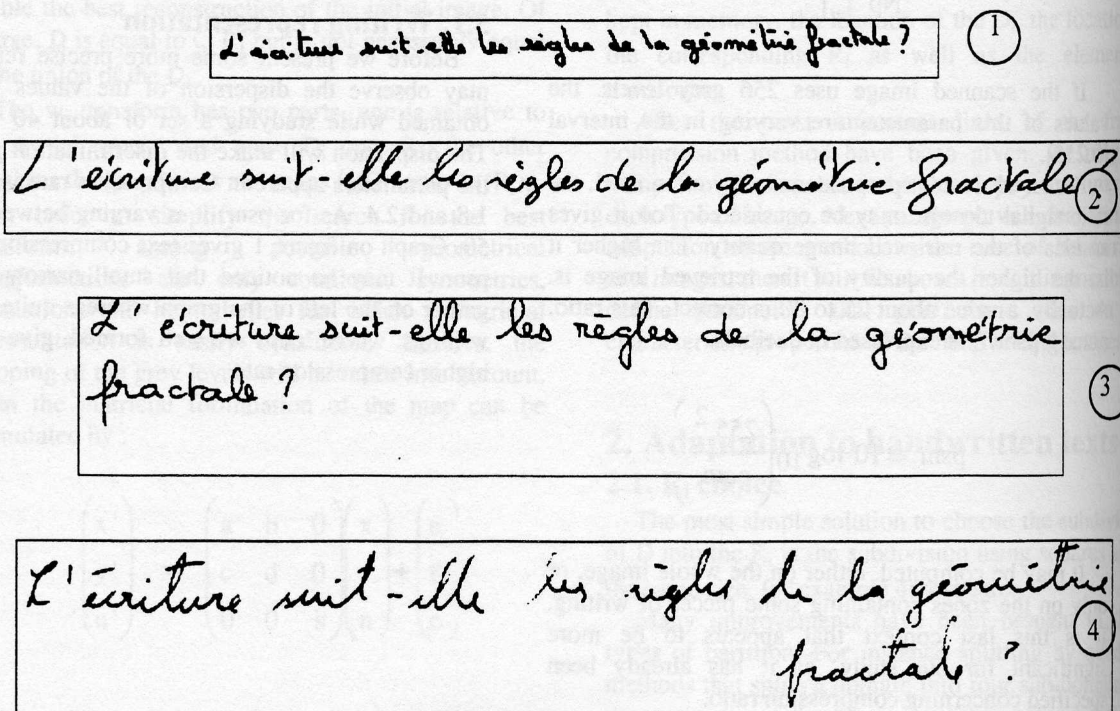

Figure 2 : Some samples of the studied handwritings

The four types of writings that have been indicated here, have been written by students in identical conditions, at the same time, and they reproduce the same sentence. Only the writers are different.

These writings have been numbered from 1 to 4. The same numbers are referred to on figure 1. It can be seen that the parameter that is used indicate from left to right a larger and larger word or global spacing (the difference between 1 and 4 is specially important).

Besides, the lisibility factor is a bit less taken into account. Nevertheless, along the x-axis, the writing number 1 that is small, not slopping and very regular, can be opposed to the writing number 4, larger and not well formed.

The spacing influence is more completely analyzed in the next paragraph.

3.2. Influence of spaces

Spaces width may vary, between letters, between words and too between the lines of a text. In all cases we observe compression rate is constant while psnr is modified, it increases when spaces are increasing. On figure 3 is shown the tendency concerning word spacing.

This modification of letter proximity is also obtained, in some sense, when writing is a forward sloping writing. We had an experiment on this case. The letter slant with respect to the writing line has been made vary with the help of an image processing software. The angles have been chosen equal to 90, 75 and 60 degrees. The same conclusions are coming out. The effect on compression rate is not significant, whereas the psnr parameter is increasing when writing slant is increasing.

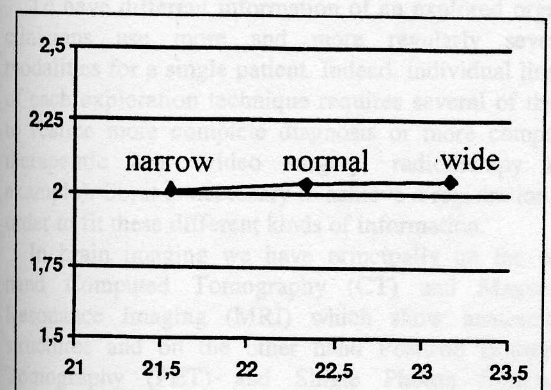

Figure 3 : Influence of word spacing

3.3. Influence of text color

Most often, for recognition task, the texts that are dealt with are binary texts. But some times the poor result may come from a poor binarisation process. So we wanted to use lost information carried by the grey level values. The grey level values may vary according to contrast chosen at the time of digitalization or according to the used pen. The more the contrast is high, the more the difference between the text and the background is important, as if the used ink color were darker. In this case the lines are specially neat and the compression/ decompression process induces a little fuzzy zone around the line of the retrieved image. This can but contribute to increase eqm and hence reduce psnr. So we have defined an index enabling to define the better contrast of the writing image. Compression rate only varies a little.

All things being equal, grey levels may be varying too when text color is changed. The experience has been carried using identical texts, written by a single writer using either blue, black, red, green or purple color ink. As far as psnr is concerned, different reactions can be seen that are linked to the used scanner characteristics. According to the experimental conditions, here we could take advantage of the result to get some information more about the used scanner than about the text itself. Nevertheless, if the texts are scanned using the same scanner, some color discrimination can be achieved. Besides, as far as compression rate is concerned, differences are not significant.

3.4. About printed texts

As a special case of writing, we were interested in testing the behavior of printed text with respect to our parameters. More specifically some texts with different fonts were concerned. Some remarks can be stated. First of all, the usual printed texts behave as the narrow writings, with more emphasize; and with the four polices that have been chosen, an easy discrimination can be achieved. The " script " police behavior differs from that of the other polices. The compression rate is higher, and psnr is much lower. Two factors are involved in these results. One is the cursive property of the script police and besides, script text may appear as being in italics. Nevertheless, if the spacing between the letters is modified in these different polices, they can be made indistinguishable.

Conclusion

It may be asserted, but this is not yet sure, the process we have just been describing, is only a reflexion of human brain activity that is processing information till it is in a form available for a specific purpose. Nevertheless, the hypothesis remains interesting.

When looking at the two parameters that have been defined, it seems the variations of the psnr would give way to a rather easy interpretation, and that this parameter would have good discriminant properties as far as some handwriting properties are concerned. Besides, compression ratio seems independent towards some characteristic properties, within the writing of a single writer ; this may have interesting consequences for the use of the parameter.

In the near future, the work is pursued by the search of some inner similarities, the nature of which is linked to the writing style.

References

[BAR88] M. BARNSLEY. Fractals Everywhere. Academic Press, Boston, 1988.

[BAR93] M. BARNSLEY, L. HURD. Fractal image compression. Wellesley, 1993, 244p.

[BOU97] V. BOULETREAU, N. VINCENT, R. SABOURIN and H. EMPTOZ. Synthetic Parameters for Handwriting Classification. ICDAR'97, August 18-20 1997, Ulm, pp. 102-106.

[CRE95] J.-P. CRETTEZ. A set of handwriting families : style recognition. ICDAR'95, August 14-16 1995, Montréal, pp. 489-494.

[FIS95] Y. FISHER. Fractal image compression, theory and application. Springer-Verlag, 1995, 341p.

[LER96] A. LEROY. Répartition des scripteurs en classes. CNED'96, 3-5 juillet 1996, Nantes, pp. 275-282.

[PLA93] R. PLAMONDON. Understanding stroke generation, a global neuromuscular approach. ICODH'93, 1993, Paris, pp. 1-3.

[PLA97] R. PLAMONDON and W. GUERFALI. The origin of the 2/3 power law. IGS'97, August 24-28 1997, Genova, Italy, pp. 17-18.

[VIN95] N. VINCENT, H. EMPTOZ. A classification of writings based on fractals. In Fractal reviews in the Natural and Applied Sciences, Chapman & Hall, 1995, pp. 320-331.