Differentiation of alphabets in handwritten texts

A. SEROPIAN¹, M. GRIMALDI¹, Pr. N. VINCENT²
¹Laboratoire d’Optique Appliquée, Université de Toulon et du Var
BP 132, 83957 La Garde Cedex – France
²Laboratoire CRIP5, Université de Paris 5, 45 Rue des Saints Pères
75006 Paris – France
E-mails : seropian@univ-tln.fr, grimaldi@univ-tln.fr, nicole.vincent@math-info.univ-paris5.fr

Abstract

Our aim is to differentiate between parts of handwritten text written using different alphabets. We achieve our goal thanks to a fractal analysis of handwriting style. For each alphabet, a set of characteristics is extracted. Advantage is taken from the autosimilarity properties that are present in the handwriting. In order to do that, some invariant patterns characterizing the writing are statistically extracted. During the training step these invariant patterns appear while performing a fractal compression process, then they are organized in a reference base that can be associated with the alphabet. The alphabet identification is performed during a Pattern Matching process using the different reference bases successively. The results of this analyze are estimated through a correlation coefficient between the initial image of the text and a synthetic reconstruction of the text based on the references.

Keywords : differentiation of alphabet, fractal compression, extraction of invariant patterns, pattern matching.

1. Introduction

Nowadays the number of documents that are created is exploding and these documents are written in many various languages using different alphabets. The automation of their treatments often relies on the identification of the alphabet in order to perform the right process. The printed documents can be processed rather well but things are getting worth as handwritten documents are concerned. Here, we tackle the problem of alphabet identification in handwritten documents [7]. It’s part of handwriting recognition domain. Knowing the language of a document allows adapting a handwriting recognition process.

Many different methods have been proposed [2,3,4], for example, Hochberg and al. [1] use connected components to extract features. They vary according to the way the alphabet is characterized. The confidence of some OCR can be used or specific shapes can be extracted or some global processing can be performed as texture analysis.

Our method relies on a statistical global approach but it also tries and extracts some specific shapes invariant in some writings rather than in others. These specific shapes are found in a learning step from a fractal decompression process applied to the handwriting image.

In the first part, the principles of fractal compression and decompression processes will be recalled. In the second part, we will show how the invariant patterns in the writing can be extracted. These invariants are learnt in order to be later used to characterize the alphabet. In the last part, we present how the similarities between writings and learned shapes are quantified to differentiate between alphabets. Some results are presented.

2. Fractal compression

2.1. Iterative function systems

Fractal compression is a technique that has been developed by Y. Fisher [5]. Its basic principle is to try and consider a given image \( I \) as the fixed point of a geometrical transform \( T \). Most often, the transform \( T \) is complex and the image is defined as the attractor of an iterative function system (IFS). The fixed point is obtained as the limit of an image sequence \( I = \lim_{n \to \infty} T(I_n) \) that is iteratively defined what ever the first term is and applying the relation \( I_{n+1} = T(I_n) \).

Then, the problem of compression is to get from a known image, a system of transforms that would precisely admit this image as its fixed point. An IFS is a set of geometrical elementary linear or affine contractive transforms that allows to generate a fractal image. These transforms make possible the definition of a function \( T \) defined by:

\[
T(I) = \bigcup_{i=1}^{n} T_i(I)
\]

The fixed point (image) that is obtained has some specific properties. In particular it is made of copies of itself but modified by the elementary transforms. In the following example, the system is made of 3 transforms, a reduction, followed by a translation for repositioning in a triangle shape. The fixed point is the Sierpinski triangle.
2.2. Fractal compression

Then, aim of compression step is to determine the transforms that are part of a PIFS having the initial image \( I \) as fixed point. To construct the PIFS, the image is partitioned in sub-images \( R \) called Ranges. These Ranges have to be interpreted as a result of a geometrical affine contractive transformation of Domain \( D \) with a \( T_i \). These Domains have to be themselves sub-images of the initial image. In order to apply fixed point theorem, in a usual way, the Domains \( D \) are twice the size of the corresponding Ranges to make sure \( T_i \) is contractive. We have: \( R = T_i(D) \). From a practical point of view, the parts \( R \) of the image with the transforms are approximated by minimizing the distance between \( R \) and \( T_i(D) \). The usual square metric is used.

![Figure 2. Principle of the fractal compression](image)

2.3. Fractal decompression

In the decompression step the transforms are iteratively applied to all sub-images of any image till the fixed point is obtained. It is assumed to be obtained when the difference between two successive images of the sequence is small enough. To quantify the quality of the fractal compression of an image, beside the compression ratio, the peak signal/noise ratio is generally used. As compression is not the goal of the process, we prefer the use of the correlation coefficient computed between two images \( F \) and \( G \), it is more representative of the visual perception.

\[
C(x,y) = \sum_i \sum_j F(i,j)G(i+x, j+y)
\]

This principle is the starting point from which the concept of our method of differentiation of alphabets is derived. Here, we will show how a base constituted from invariant elements can be built for each alphabet.

3. Alphabet style learning

In this section, the possibility to extract the invariant elements from handwriting [6] is demonstrated. Each alphabet is considered as the combination of invariant shapes that are modified by each writer. We consider a handwritten text image as the aggregation of letters. Our method is relying on the inner similarities that still exist in the handwriting in spite of the deformations. These elements are extracted during the fractal compression process that is to say while searching for the transforms of the PIFS.

During the compression step, the objective is that all the Ranges can be obtained as transformation of a Domain of the image. The criterion to choose the best transform is the RMS (Root Mean Square) between the two \( R \) and \( T_i(D) \) where we make \( D \) vary. The Domains are chosen among any sub-images of appropriated size contained in the image.

In fact, we are not interested in the transforms but in the subimages (Domains) that are selected. They contain elements that can be seen at least two different scales. Thus, for documents using a same alphabet, a reference base \( B \) is built from the set of Domains that have been extracted and used in the compression selection. These domains contain some characteristics of this way of writing. In order to be significant of the alphabet, the text used in the learning phase has to be long enough and to be
written by as many writers as possible to represent all varied writing styles. The reference base represents the inner similarities contained in the writing.

When looking for the possible transforms to best associate Ranges and Domains, we have limited the search to only one type of transforms: the homothetic transform with ratio 1/2 associated with a translation. Besides, the choice of a partition of the image induces the geometrical shape of the sub-images that constitute the Ranges and the Domains. At this stage we have chosen square windows and a spatial contraction ratio of $\frac{1}{2}$.

4. Differentiation of alphabets

Of course, in order to differentiate an alphabet from another, the learning phase concerning each alphabet must have occurred and the corresponding reference bases must have been stored. If N alphabets have to be differentiated, a priori we have got N reference bases.

The comparison between handwritings is only possible when the sizes of the writings are identical. So, it’s necessary to apply a normalization step on the writings. As a matter of fact, the height is linked to the choice of the dimensions for the fixed size of the range sub-images and has to be adapted to the usual details of the writing. They may differ from one way of writing to another as the alphabets are not the same. The writings are normalized with respect to the height of the letters. More precisely, a scalar central symmetry is computed according to the vertical direction and applied according to both directions on the image. The most important is to apply the same process to every text that is studied.

The length of the text to be studied is not necessarily very long compared to the learning text. In any case it usually contains only one writer style. The text will be processed.

Likewise in the fractal compression, the image is partitioned in the same way in Ranges. Then, no inner similarity is looked for, but a pattern matching process between the ranges of the text to be identified and the elements of the reference bases is performed. The ranges have the same shape as the sub-images of reference bases. For each reference base and in order to quantify the correspondence between the text and the reference base, a new image is built from a white image where are copied the sub-images when the correlation coefficients with the ranges are large enough. The quality of the correspondence is measured by comparing the initial text and the reconstruction of the image that is achieved by a pattern matching process (cf. Figure 3).

Thus, with each Range $R_i$ of the non identified image, a sub-image of the base $B_j$ is associated when minimizing the square error criterion $\sum (R_i - T(D_j))^2$. Here again, the transform considered is based on the homothetic transform with a ratio of $\frac{1}{2}$.

5. Results

We have considered two learning texts written in two different languages. We have taken some French and Arabic texts. Both texts are built from samples of handwritings by different writers. In each case we have considered 10 different writers. Some of the samples are shown on Figure 4.

After the learning step, we obtained two bases (cf. Figure 5) from these two texts. Figure 5(a) represents the Arabic base with 365 sub-images and Figure 5(b) the French base with 279 sub-images. The differentiation method was then tested on different texts. We obtained Table 1. Indeed, a text on test is more or less similar to the base used. In each column are indicated the average values of the correlation coefficients obtained on the reference base. All texts have been well sorted and we see the difference while using either latin or arabic reference bases is significant enough to perform the differentiation between alphabets.

(a) part of Arabic learning text, 3 different writers

(b) part of French learning text, 3 different writers

Figure 4. Learning texts for two alphabets

Figure 3. Differentiation step

The text will be assigned with respect to the reference that allows the greatest number of correspondences. The quality of the similarity between the initial image and the synthesized image is quantified using the correlation coefficient parameter.
6. Conclusion

Here, we have shown that alphabet differentiation can be achieved prior to any recognition task as a texture differentiation. But we have not used image processing techniques that are too low level oriented. The features we have extracted are characteristic of the alphabet and rather high level information is contained in the elementary elements extracted. The response we have given relies on an original approach. The image is treated as a whole taking into account specific characteristics of each alphabet, without neither any specific a priori pattern extraction nor details comparison. The results are convincing, of course they have to be confirmed by more experimentations and with other alphabets as Cyrillic and Asian ones.

7. References