A multiscale method for text and line extraction from gray-level noisy images

Thierry Frêche and Nicole Vincent
Laboratoire d'Informatique
Université de Tours (LI/E3i)
64, avenue J. Portalis
37200 Tours, France
{freche, vincent}@univ-tours.fr

Abstract

The goal of our study is noisy background removal in a text document image. Either a classical binarization can be achieved in order to use an OCR system, or the extraction of interesting zones containing the written lines.

For the problem of text extraction from any document, in spite of the good results in some particular situations, no general solution has been found yet. Here we propose a method that can be applied to a wide range of problems.

Of course, our method is relying on the use of the gray levels. The gray levels of the text images are interpreted as levels or altitudes in a geographical topographical map. Our approach is semi-local. It defines some influence zones associated with focused zones. The process comprises the study of the evolution of one parameter, at different scales. This can be seen as the evolution of human perception of the document according to observation scale. The interest of the method lies in the study of the evolution of a phenomenon rather than on a static fact. An optimum value is computed, that has some perceptive meaning and the computation of that is made easier because of the monotonous property of the function to be studied.

Significant results, according to these techniques, have been obtained on various examples and some different application domains can be addressed with this method.

Keywords

segmentation, local processing, multiscale approach, gray level documents, text extraction

Introduction

The extraction of data from noisy background is a difficult problem. The large field of image segmentation is concerned. Many technical solutions have been proposed and are still under development. In the field of image processing, the most commonly used method is segmentation or binarization [8, 9, 10, 11, 2, 6]. Generally, the methods developed to achieve this separation between noise and the zones of interest in the images can be classified into three large families.

One is composed of the methods based on edge detection. Those methods consist in finding the local discontinuities of the signal. Then, these singularities are organized in order to get connected shapes [5, 3]. The second family includes all the techniques looking for some adjacent regions with the same property [7]. Some can be achieved at different scale levels and so can be more or less coarse or precise. For example, some criterion can be used to find the homogeneous statistical properties of the background. Starting from a global vision of the image, the criterion can adapt itself to local properties. The third family is built from mathematical methods of filtering [4, 1].

These different methods give good results when they are applied to the problems for which they have been developed. But when some parameters of the images vary, the domain or the contrasts for instance, they lead to less satisfying results. This is specially the case when bank checks are concerned. The quality of the result is depending on the complexity of the background and on the colors involved.

In this paper we concentrate on the study of gray level documents from which text parts are to be extracted. This problem occurs for example when bank checks are processed or when envelopes are sorted by postal services. The knowledge of the gray levels gives information. This information is very important to segment the handwritten words correctly.

An other interesting application of this task is image compression, only informative parts being left in the document after processing.

After we have presented our motivation to introduce a new approach to text and line extraction from images, in a second part we explain the different steps and philosophy of our method. In the third part we show how the principle leads to text and line extraction. In the last part some results are shown, and we can see how the method can be applied to other segmentation problems.

1. Why a new approach?

In order to process text images and to take advantage of the gray level images it is often necessary to first take out the non-significant parts. Those that can be considered as noise. This action can be seen as a segmentation step. The lines are extracted from the background. The criterion is the darkness of the significant parts. As no universal solution has been yet found for segmentation purpose, we wanted to take into account the particularities of text images, particularly the two-class problem (text and background) and the importance of the contrast between the text and the background to make the text easy to be read. The usually used methods give some efficient results in particular contexts. Let us consider the common characteristics of most approaches, and see how is emerging our method.

1.1. Local versus global approaches?

Our human vision of any problem starts from a global vision of an image. It becomes local when human is looking for details that could give some information and help building a conclusion without the need to achieve a complete analysis. Nevertheless, the use of a global model, most often, leads to non-satisfying results. A global approach cannot give enough information about the dark pixels that are present in an image. On the whole the black pixels are statistically less numerous. Besides, a global image has rarely some homogeneous properties. The contrast is varying all along the page, depending on the origin of the light or on the presence of different folds in the sheet of paper, even on the different pens that have been possibly used, on a background image that can be present.

In fact the human decisions are to be taken locally, after the significant context is determined. Nevertheless, an early global vision is necessary. The automatic methods have to be able to adapt to local specificity of the image.

When using a histogram, or any other function, a critical value is looked for that separates two domains. A global value, valid on the whole image, does not take into account local properties either of the lines or of the background. In most methods, a local window on which the process is applied figures the local context.

One of the difficulties with these methods is a block like aspect that may appear in the final image resulting from the non-continuous process on the border of each window. Too, an other one is the choice of the right size of the window in which the study is done. The good sizes are too domain dependant. For text extraction, the size will depend on the size of the character body.

A multiscale approach seems quite interesting in order to cope with the duality between local and global vision.

1.2. The optimization problem

The core of most methods, what ever the domain concerned is, is to find an optimum. This optimum may be found in the gray level histogram or from a gray level function defining the image itself, or from an energy or entropy function.

Indeed, the non-significant local extremes are difficult to differentiate from global or significant extremes.

2. A multiscale approach

For us, this way appears to be quite necessary and brings an essential character to our method.

2.1. The characteristics of the method

Let us emphasize here some constraints that must be answered by the method we are to elaborate.

- We are to define a monotonic function from the gray level image, from which a critical value will allow in a first step to separate the background from the significant part.

  For mathematical reasons, it is more convenient to work from monotonic functions.

- The method must take model on the human vision. That is to say create a context and adapt to it.

- Different observation scales figure the context, that is to say the influence zone around the focused zone.

The local zones are windows overlapping, pixel by pixel, but rather they are considered as neighborhoods of a focused zone. We model the human vision when the observation of a point takes implicitly into account the environment, the context. The conclusion will not be valid on the whole zone that is integrated, but only associated with the focused zone.

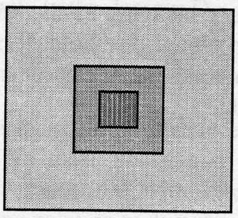


Figure 1: focalisation zone and influence zone

- It is not safe to conclude from only one situation, only one context or observation domain. Rather than looking for the best scale in order to deduce the threshold value, all the values computed must be taken into account.
- When are to be extracted significant dark zones in the image, it must be reminded that, in fact, the actual image is obviously not a binarized image, but rather we can imagine the image, the pixel gray level, has been altered and the original contrast, that should have been, is diminished. So when hypothesizing a black zone, the gray black levels are to be enhanced in order to compute some quantitative parameter values.

2.2. From local to multiscale vision

The consequences due to the choice of the size and shape of an observation window have to be minimized. Besides in order to model human way of observation, with each zone are associated neighbors of the zone, many sizes are considered. In some way it is a modification of the observation scale as a larger zone is concerned for the perception of a fixed zone.

Nevertheless, the final decision will only concern the focused zone. So the problem linked to non-overlapping zones is solved. The management of the discontinuities occurring on the outline of the zone is solved too. Here the results applied to one zone are depending on a larger zone, each pixel is considered in several zones, its gray level value is influencing several focalization zones. The different neighborhoods of a focalization zone can be built as their dilated sets, by means of morphological transforms. For each observation scale, a threshold can be computed that is applied in the focalization zone.

2.3. A dynamic decision

The choice of the final threshold is not obtained through an explicit determination of the value giving an optimum to an optimality criterion.

Here we are taking into account all the threshold values that have been computed in one influence zone or another. The global decision is obtained through the fusion of the different results. We want to privilege the value that is the most frequent among all the results. This plan is modified in order never to neglect the result obtained in the most local way with respect to the focused zone. For this no weight has been defined associated to each neighbor, but rather starting from the smallest zone, larger and larger zones are considered till a significant majority is emerging.

3. Adaptation of the method

3.1 Image interpretation

Then a study zone D is considered and its influence zone I. For instance I is the transformed image of D through a morphological dilation operator. All the pixels in I are to be taken into account to find a threshold suited to the I context, but the result is intended to be applied only on the D domain.

A parameter has to be chosen. The evolution of its value according to a potential threshold value will allow defining the best threshold value. Here a volume is the parameter to be computed. Instead of looking at the image as a matrix of gray level pixels, let us imagine the gray levels are indicating the altitude of the points of a 3D-surface. Then a thresholding of the image only lets the highest points remain from the image. Each point contributes to the volume through a vertical column between the threshold value and the upper surface. The volumes of such landscapes are to be computed. More precisely, the evolution of the contribution of the I region to the volume will be studied when the threshold is evolving.

3.2. Gray level scale

It would be obvious to consider the 255 possible volumes associated to a 256 gray level image. But this would have several disadvantages.

First of all, human has no need of 255 gray levels to be sensible to nuances within a text image. Further more human eye is not able to distinguish between nearby such values. The number of different values can be more reasonable.

Besides, taking into account all the gray levels would lead to too important computation that would increase process time. From an other point of view different zones will be compared and the contrast on each of them will be different. In order to make comparisons possible, normalization has been performed. On each zone, the gray levels are comprised between a minimum and a maximum value, let us say Min and Max, what ever the absolute value be. On this interval, the amplitude of which varies, the reference points have to be fixed with respect to the extreme values.

In order to solve the competition between computation speed and result precision we have decided to consider only N reference values that are uniformly distributed between Min and Max. In the experiment 10 values are considered.

$$R_k = Min + k\,(Max - Min)/N \qquad k \in [0, N]$$

We cannot forget that the shapes, the text we want to extract have been "eroded" and that we want to restore the dark portions. Then in order to increase the actual difference between the background and the potential text, we replace in the volume computation the altitude of a pixel zone by the Max value as far as the altitude is superior to the considered reference value. This is illustrated in figure 2.

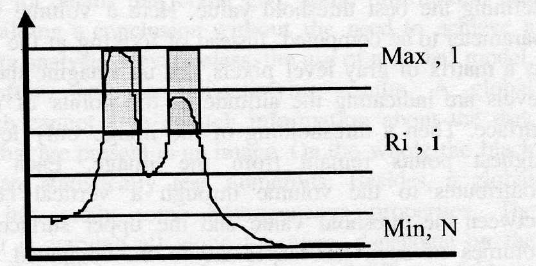


Figure 2: modified gray levels of the upper parts

The volume associated with reference R_k is defined by:

$$V_k = \sum_{x \in I} s_k(f(x)) * (Max - R_k)$$

where f notes the gray level function of the image and s_k is a function that is equal to zero if its argument is less than R_k and is equal to 1 otherwise.

3.3. Threshold value in a fixe area

In such a way, N volumes can be computed. By definition, these values are increasing with respect to the increasing index. The unit of the volume is the voxel. An example can be seen in figure 3, the volume is evolving with respect to a potential threshold. It can be noticed that the x-coordinate indicates the reference rank and no absolute value is considered.

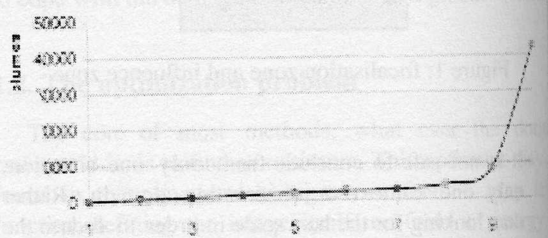


Figure 3: volumes associated with an influence zone according to the reference value order

Now let us give an explanation of this representation.

In the most right part, the increase is explained because all the values of the noisy background pixels are summarized.

On the left, only the dark pixels of the line of writing are involved and the curve is increasing slowly.

Then the way to find a good threshold has to be defined. If the noisy pixels are those that make the curve rise up, the threshold is the value corresponding to the moment the curve takes off.

If the image had no line drawn on it, and all the gray levels between Max and Min would be uniformly used, the graph would appear with a linear aspect. Then the essential point is a rupture point in the graph. It has been computed from the first step of a polygonalisation of the curve. The rupture point is the furthest point of the curve from the straight line joining the extreme points. The abscissa of this rupture point gives a threshold value. For a rough approximation, we can consider here the nearest higher integer value. When no text is present on the local zone, the line will be horizontal.

Once the value of the corresponding threshold is found, it is easy to get the numerical correspondence in the scale of the gray levels. So the enhancement of dark elements could be performed.

But before this step we shall achieve the same process for different influence zones the size of which is made vary. With each influence zone I_n is associated a threshold T_n.

3.4 Final threshold value

Now comes the last step to achieve the decision of the best threshold to apply to the domain D.

p thresholds are computed according to the different influence zones.

In figure 4 are indicated the different evolutions of the volume parameter. To make possible a comparison, the volumes have been normalized with respect to the first reference. The rupture point on each curve gives a potential threshold value.

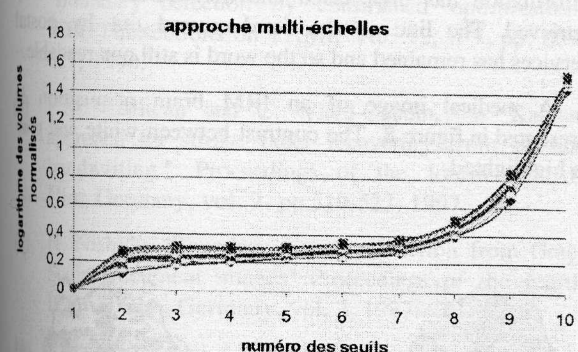


Figure 4: Study in different influence zones of a focalization zone

For the final value, we choose the most frequent occurrence. More precisely, the set of threshold values is organized in M classes. The classes have all the same amplitude and the most frequent class is selected. A value is deduced from this class and the adjacent ones. This is the basic idea.

In order to improve the rapidity of the process each time a threshold value is computed, the possibility of a decision is looked for. The order the influence zones are studied is linked to the increasing dimension of the zone.

After the n^{th} calculation, the decision can be taken when

$$T_n \in Cl_i \text{ and } \forall j \in [1,N] \setminus \{i\} \quad |Cl(T_n)| > |Cl_j| + d$$

where Cl(T) is the class of the T value.

Our experiments proved a d value equal to 10 leads to good results.

When the final value is computed from the distribution of the T values, let us say T_i, we have to recover a threshold value rather than an order value. The conversion is done by use of the scale that is associated with the i^{th} study of influence zones.

In next part some examples of the processed images are presented.

4. Experimental results

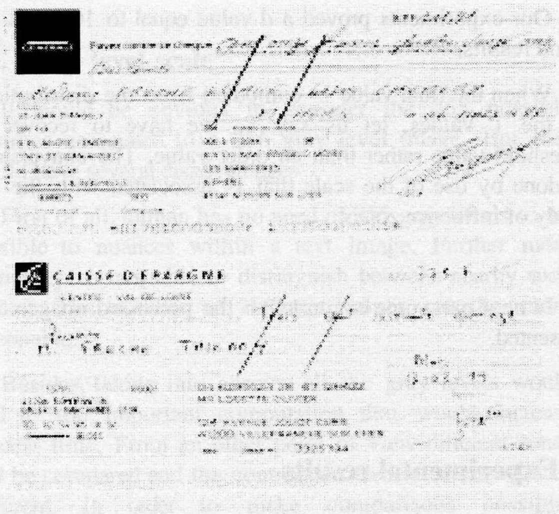
This process has been applied to many text images, and too some medical images have been studied.

The results of these trials have led to images with increased contrast or to a mask such that information within the text lines that have been detected, has not been lost.

The first experiment is dealing with check images. The first one is rather dark with a background predominantly dark red. On the resulting image that can be seen in figure 5, the text has been well recovered and the handwritten indications are quite neat.

The second check image, too in image 5, has a background comprising various designs with gray colors, predominantly anthracite. The handwritten indications are also well extracted. The logo in upper left corner has even improved.

(a) initial image



(b) processed images

Figure 5: Example of bank cheks images and processed images

A second type of example has been chosen in particularly noisy text images. In figure 6 the first number is very noisy and the background presents non uniform properties. Nevertheless the number on the processed image is quite readible. The method shows here its limit and a filtering could be added in order to minimize the remaining background noise.

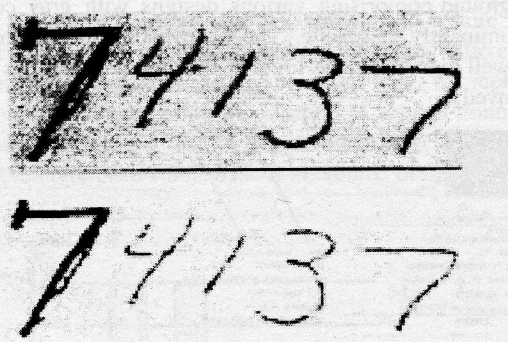
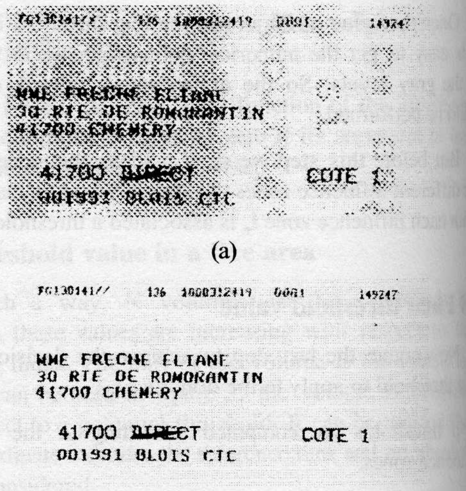


Figure 6: Example of noisy handwritten numerals

In next example presented in figure 7, the quality document is not better, but has quite different origin. An address printed on a plastic content is concerned and the underneath information is quite visible on the document.



(a)

(b)

Figure 7: Example of a plastic document

On the processed resulting image the background information has disappeared and the address is perfectly retrieved. The line on the word crossed out by postal services has remained and so the word is still non readable.

A medical image of an IRM brain acquisition is presented in figure 8. The contrast between white and gray is highlighted.

Figure 8: Brain image

This gives an example of the wide range of possibilities that the method can offer.

Conclusion

This study has presented an original method, using a multiscale approach and a quite local approach to adapt to the specificity of the areas to be processed.

Using these techniques, significant results have been obtained when there is a low contrast. The technique finds and enhances the small details with respect to the difference of contrast in a noisy gray level image.

In the future, the combination of the several methods will be tried: such as enhancement by volume segmentation, morphological filtering and then isolation of the text zones by variance discrimination at a small scale in order to improve this new technique.

References

[1] M. Cheriet, " Extraction of handwritten data from noisy gray level images using a multi-scale approach " proceedings of Vision Interface, pp. 389-396, 1998.

[2] O. Deforges, P. Piquin, D. Viard-Gaudin, D. Barba, "Segmentation d'images de documents par une approche multirésolution, extraction précise des lignes de texte" Traitement du Signal, Vol 12, n°6, 1995, pp. 527-539.

[3] S. German, D. German, C. Graffigne, P. Dong, "Boundary detection by constrained optimization" IEEE Transactions on PAMI, Vol. 12, n°7, pp. 609-627, 1990.

[4] P.M. Lallican and C. Viard-Gaudin, " a Kalman approach for stroke order recovering from off-line handwriting " Proceedings of the fourth ICDAR Ulm, Germany, vol. 2, pp.519-522, 1997.

[5] H. Nishida, "Boundary feature extraction from Gray-Scale Document images" Proceedings of the fourth ICDAR Ulm, Germany, vol. 2, 1997.

[6] R. Pal and S. Pal, " A review on image segmentation techniques " Pattern Recognition, vol. 26, n°9, pp.1277-1294, 1993.

[7] S.S. Reddi, S.F. Rudin and H.R. Keshavan, " An optimal multiple threshold scheme for image segmentation " IEEE transaction on S.M.C., vol. 14, no.4, pp.661-665, 1994.

[8] P.K. Sahoo, S. Soltani, A.K.C. Wong, and Y.C. Chen, " A survey of thresholding techniques ", computer vision, graphics and image processing, vol. 41 pp. 233-260, 1980.

[9] O.D. Trier and A.K. Jain, " Evaluation of binarisation methods for document images " IEEE trans. On PAMI, vol. 17, n°3, pp. 312-315, 1995.

[11] L. Wang and T Palvlidis, "Direct gray-scale extraction of features for character recognition" IEEE transaction on PAMI, vol. 15, n°10, pp.1053-1067, 1993.