

ROBUST THRESHOLDING BASED ON WAVELETS AND THINNING ALGORITHMS FOR DEGRADED CAMERA IMAGES

Céline Thillou and Bernard Gosselin

celine.thillou@tcts.fpms.ac.be
Faculté Polytechnique de Mons, Avenue Copernic, 7000 Mons, Belgium

ABSTRACT

This paper describes a thresholding method for degraded documents acquired from a low-resolution camera. This technique is based on wavelet denoising and global thresholding for non-uniform illumination. The accent is put on stroke analysis to keep useful information more accurately without breaking characters and by reducing the number of overconnected ones. An improvement of the technique, which uses the thinning algorithm, is thus detailed and a particular attention is given to use this optimization in the convenient images only. Moreover, thanks to the wavelet decomposition, complex backgrounds as well as high frequency noise can be considered. This method can handle various corpus of images without need of any prior knowledge of the document image and fine-tuning of parameters.

Keywords : Wavelet, thresholding, denoising, degraded documents

1. INTRODUCTION

1.1. Context of Images

As the first step of an entire system for optical character recognition, thresholding is crucial and its success is preponderant for all other processings. Errors at this point are propagated all along the recognition system. The challenge to obtain a very robust thresholding method is major. Actually, after thresholding, several steps are needed such as skew correction, line and word segmentation, character segmentation and finally character recognition. If noise is very present, line or word segmentation could be degraded. In a more intensive way, for character segmentation, letters can be broken or they can touch because of a bad thresholding of the document.

In our context, documents are acquired by a digital camera (1280*1000 pixels) and taken by blind or visually impaired people [5]. All these constraints lead to several kinds of degradations and especially several irregular or non uniform degradations, which could not be forecast. We have no a priori information on pictures and no validation from the user.

Therefore, this context presents a bunch of degradations, missing in classical scanner-based pictures, such as blur, perspective distortion, complex layout, interaction of the

content and background, uneven lighting, wide-angle lens distortion, focusing, moving targets, sensor noise, intensity and color quantization.

Some of these degradations could be corrected thanks to a robust thresholding which improves the character recognition. Moreover it minimizes the other degradations which become useless to solve.

Therefore, our main objective is to provide an automatic thresholding method which improves drastically character recognition and which is as general as possible to handle a large and various corpus of images, as those which could be taken by a blind user.

1.2. A Brief State of the Art

Two modes are usually used for thresholding. The first one, global thresholding, finds a single threshold for all image pixels. These methods are mainly general but not robust. Actually, it is very fast and efficient when the pixel values of the components and those of the background are fairly consistent. The other one, local thresholding, uses different threshold values which are required for different local areas. These methods are more robust but not as general because of many parameters to tune.

Although many thresholding techniques [15], [7], such as global [11] and local [8] ones, have been developed in the past, it is still difficult to deal with images with very low quality. Meanwhile, in our context, or in general, documents processing systems need to process a large number of documents with different styles, thus, they require that whole processing procedure is achieved automatically and without prior knowledge and pre-specified parameters.

With similar approaches as ours, Sauvola [12] and Chan [2] processed only uneven lighting in time-domain. Yang [16] developed a thresholding method based on gradient with edge preservation on poor quality documents without complex backgrounds. Seeger [13] created a new thresholding technique for camera images, like in our context, by computing a surface of background intensities and by performing adaptive thresholding for simple backgrounds. In wavelet or frequency domains, Zhang [17] and Liu [9] considered bright objects but not characters, therefore wi-



FIG. 1 – From top to bottom : the original image, the wavelet decomposition with multiple subimages, the thresholding on the initial image and the thresholding on the reconstructed image

thout stroke analysis. Dawoud [4] used wavelet reconstruction based on an image dependent model for cheques. Berkner [1] had a technique on wavelet-based denoising, sharpening, and smoothing in Besov spaces but no compromise is done on stroke analysis.

In this paper, we will present our technique based on wavelet denoising and Otsu [10] thresholding for non-uniform illumination. A particular interest is given on stroke analysis to reduce the number of touching characters and to prevent broken ones. Moreover complex backgrounds as well as salt and pepper noise are also considered thanks to the wavelet decomposition.

2. THE THRESHOLDING APPROACH

2.1. Preprocessing

First of all, we apply a contrast-enhancement preprocessing step by using *top-hat and bottom-hat filtering*. The top-hat transform is calculated by subtracting the result of morphologically opening the image from the original image and the bottom-hat one by subtracting the original image from the result of morphologically closing the image.

2.2. Problem of Non Uniform Illumination

An important problem for thresholding methods comes from a *non-uniform illumination* which introduces most of the noise when using only a global method. This bad illumination appears as wide noisy areas, so we assumed that the illumination noise has a lower frequency spectrum than the character one.

We decided to apply an *orthogonal wavelet decomposition* of the initial image. The wavelet transform splits the frequency range of the image into equal blocks and represents the spatial image for each frequency block. It offers a natural framework for providing multiscale image representations that can be separately analyzed. For example, through a multiscale decomposition, most of the gross intensity distribution can be isolated in a large scale image, while the information about details and singularities, such as edges, can be isolated in mid- to small scales. We use a level 8 wavelet transform using the *Daubechies* [3] 8 wavelet.

We see very well in the LF7 and LF8 images of Figure 1 the low frequency noise due to the bad illumination. Our idea was to eliminate all low frequency wavelets except the first one which is called LF1 and contains anyway the main useful information.



FIG. 2 – Results with the improvement of the algorithm. From top to bottom : the thresholded reconstructed image, the mask and the skeleton on the image and the result with the new threshold (the mean gray level values)

The final reconstructed image will be :

$$R = LF1 + HF1 + HF2 + HF3 + HF4 + HF5 + HF6 + HF7 + HF8$$

Then we use the well-known Otsu [10] method directly on the reconstructed image and we obtain the binary image. In Figure 1, we can see the difference between the Otsu method applied directly on the initial image and on the reconstructed image.

2.3. Problem of Touching Characters

Sometimes for text with thick characters, a few ones touch each other and must be split to minimize the number of touching characters.

The main idea is to pick the mean gray-level value of characters. Actually, the gray-level value of a character is homogeneous in the center of the character at least. For blurred or thick characters which become touching ones, the border of the character has not exactly the same gray-level value as the center.

We use the skeleton of the thresholded reconstructed image. The skeleton is the result of the thinning, a morphological operation, in which binary-valued image regions are reduced to lines that approximate the center lines of the regions, as shown in Figure 2 in the middle inside characters.

The basic iterative thinning operation [6] is to examine each pixel in an image within the context of its neighborhood region of at least 3x3 pixels and to “peel” the region boundaries, one pixel layer at a time, until the regions have been reduced to thin lines, with a width of one pixel.

A mask with ON-pixels of the binarized reconstructed image is used in order to keep the noise reduction done by the wavelet decomposition. The mask is created by multiplying the initial image and the inverse of the thresholded one. Then, all gray-level values of the skeleton are picked

out to compute the mean which becomes the new global threshold. Then a simple global thresholding can be computed on the mask.

Nevertheless, this algorithm can be applied only with thick characters to avoid to break thin characters after thresholding. For that reason, we need an algorithm to automatically detect thick characters in a text. As the skeleton is already defined, we compute the ratio between the area of connected components and the area of skeleton to test the thickness and to apply this improvement or not.

2.4. Problem of Complex Backgrounds

The *background texture* or image has some details of quite the same size and intensity as the characters. As this noise is in the same frequency domain as characters, it cannot be removed directly by wavelet decomposition. An example of this type of noise is given in Figure 3 where the background apples have quite the same characteristics as characters.

An iterative thresholding algorithm, using the Otsu method again, is thus applied directly on the previous mask and gives better results as we can see on the last picture in Figure 3. Actually, the mask is composed of two main gray level classes, the character and the noise. These two classes can be easily separated by another Otsu thresholding. A complex background does not need to be detected because when just a simple background is present, it removes only the blurred part around characters, which becomes the second class.

2.5. Salt and Pepper Noise

Another problem, very usual in camera-based images, is due to high frequency salt and pepper noise, appearing in the HF1 subimage in Figure 1, which can therefore be removed. But as HF1 can contain useful information when this kind of noise is absent, a detection solution is neces-



FIG. 3 – Improvement by re-applying the Otsu thresholding. From top to bottom : the reconstructed image, the thresholding on the reconstructed image and the iterative thresholding on the mask

sary by counting the number of very small components. This type of noise can be seen in Figure 4 and we can see, in the two images in the middle that using the wavelet decomposition already performs the thresholding. Nevertheless, on the last picture, the result is better without HF1 subimage. A low-pass filter could also be used but wavelet decomposition is already performed so the overall computational time is reduced. Moreover, this kind of filter needs additional parameters (mask size, frequency) and our goal is to remain as general as possible.

3. RESULTS, CONCLUSION AND FUTURE WORK

Our method is able to threshold various degraded document images without need of any prior knowledge of the document images or manual fine-tuning of parameters.

Several tests were done on the public ICDAR 2003 database, which is composed of camera-based pictures and some of them are bad-illuminated. It is very difficult to know what degree of uneven illumination there is in ICDAR images. Nevertheless, on the whole database, based on visual judgement, 7% images have been improved for character segmentation and recognition. We have also realized tests on a personal database with different kinds of illumination and the improvement is around 10%.

Moreover this algorithm keeps useful information more accurately without breaking strokes of the characters and by minimizing the number of touching ones. As an improvement of this method, we could use a more specific thresholding method than the Otsu one after wavelet decomposition. Moreover, more complex backgrounds need to be tested to reinforce this method in the particular case of decorative backgrounds.

4. ACKNOWLEDGEMENTS

This work is part of the project Sypole [14] and is funded by Ministère de la Région wallonne in Belgium.

5. REFERENCES

- [1] K.Berkner : Enhancement of scanned documents in Besov spaces using wavelet domain representations, *Proceedings of SPIE*, **4670** (2002) 143–154
- [2] F.H.Y.Chan, F.K.Lam and H.Zhu : Adaptive thresholding by variational method, *IEEE Transactions on Image Processing*, **7** (1998) 468–473
- [3] I.Daubechies : Ten lectures on wavelets, SIAM (1992)
- [4] A.Dawoud and M.Kamel : Binarization of documents images using image dependent model, *ICDAR 2001*, (2001) 49–53
- [5] S.Ferreira, C.Thillou and B.Gosselin : From picture to speech : an innovative application for embedded environment, *ProRISC Conference* (2003)
- [6] R.C.Gonzales and R.E.Woods : *Digital Image Processing*, Second Edition, NJ : Prentice-Hall (2002)
- [7] A.K.Jain : *Fundamentals of digital image processing*, Englewood Cliffs, NJ : Prentice-Hall (1986)
- [8] J.Kittler, J.Illingworth : Threshold selection based on a simple image statistic, *CVGIP*, **30** (1985) 125–147
- [9] J.Liu and P.Moulin : Image denoising based on scale-space mixture modeling of wavelet coefficients, *ICIP 1999*, (1999)
- [10] N.Otsu : A thresholding selection method from gray-level histogram, *IEEE Transactions on Systems, Man, and Cybernetics*, **9** (1979) 62–66



FIG. 4 – To remove salt and pepper noise. From top to bottom : the initial image, the binarized initial image, the binarized reconstructed image and the same one without HF1

- [11] P.K.Sahoo, S.Soltani, A.K.C.Wong : A survey of thresholding technique, *CVGIP*, **41** (1988) 233–260
- [12] J.Sauvola and M.Pietikäinen : Adaptive document image binarization, *Pattern Recognition*, **33** (2000) 225–236
- [13] M.Seeger and C.Dance : Binarising camera images for OCR, *ICDAR 2001*, (2001) 54–59
- [14] Sypole project : Retrieved April 9, 2004 from <http://tcts.fpms.ac.be/projects/sypole/sypole.html>
- [15] Ø.Trier, A.K.Jain and T.M.Taxt : Feature extraction methods for character recognition - a survey, *Pattern Recognition*, **29** (1996) 641–662
- [16] Y.Yang and H.Yan : An adaptive logical method for binarization of degraded document images, *Pattern Recognition*, **33** (2000) 787–807
- [17] X.P.Zhang and M.Desai : Segmentation of bright targets using wavelets and adaptive thresholding, *IEEE Transactions on Image Processing*, **10** (2001) 1020–1030