SEGMENTATION-BASED BINARIZATION FOR COLOR DEGRADED IMAGES

Celine Thillou, Bernard Gosselin
Faculte Polytechnique de Mons, Belgium
celine.thillou@tcts.fpms.ac.be

Abstract Recently, a new kind of images taken by a camera in a "real-world" environment appeared. It implies different strong degradations missing in scanner-based pictures and the presence of complex backgrounds. In order to segment text more properly as possible, a new binarization technique is proposed using color information. This information is used at proper moments in the processing not from the beginning to have smaller regions-of-interest. It presents the advantage of reducing the computation time. In this paper, an accent is put on stroke analysis and character segmentation. The binarization method takes it into account in order to improve character segmentation and recognition afterwards.

Keywords: Binarization, Color Clustering, Wavelet, Character Segmentation

1. Introduction

A new kind of camera-based images in a mobile environment appeared very recently. They can be taken by an embedded camera on a personal digital assistant or other mobile devices. Therefore this context implies a bunch of degradations, not present in classical scanner-based pictures, such as blur, perspective distortion, complex backgrounds, uneven lighting...

Thresholding, as the first step of OCR, is crucial. Actually this is the first step where some information is lost after picture acquisition. Errors at this point are propagated all along the recognition system. The challenge to obtain a very robust binarization method is major.

In our context, text areas are already detected and therefore this part is no more considered here. Our test database is based on public images given in the ICDAR 2003 website, used for their robust reading competition. Unfortunately, from this competition, no paper was written on the text binarization and recognition, only on text detection.

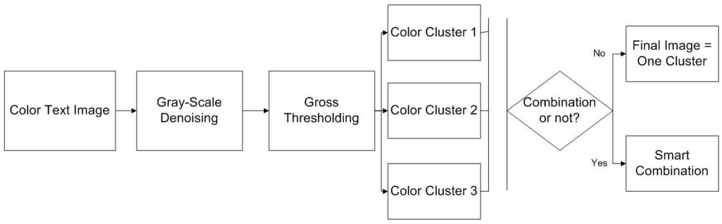

Figure 1. An overview of our binarization method

A Brief State of the Art

Most existing binarization techniques are thresholding related. Basically, these techniques can be categorized into two categories : global [6] and local or adaptive [3]. Global methods tempt to binarize the image with a single threshold. By contrast, local methods change the threshold dynamically over the image according to local information. Meanwhile in our context, image processing systems need to process a large number of documents with different styles and without pre-specified parameters. Moreover, all these techniques perform poorly under complex backgrounds.

In [4], Liu and Srihari used the global Otsu [5] algorithm to obtain candidate thresholds. Then, texture features were measured from each thresholded image, based on which the best threshold was picked. Color information is not used and this technique fails when different colors with almost the same intensity are present. Seeger [7] created a new threshold technique for camera images, like in our context, by computing a surface of background intensities and by performing adaptive thresholding for simple backgrounds.

Wang [9] tried to combine both color and texture information to improve results. This technique works well for images similar to our database but computation time required is very high and no consideration on connectivity between components are presented. With other techniques, and some similar ones, our method fills these failures.

Some great results are also obtained in the domain of content retrieval and video segmentation in multimedia documents. Garcia [2] uses color clustering for binarization. This method is based on k-means and decision about which cluster or combination of clusters has to be considered is based on a bunch of criteria concerning characters properties.

2. Our Binarization Approach

A scheme of our proposed system is presented in Figure 1. Color information is only used after gray-scale denoising and coarse thresholding in order to consider only useful parts and to decrease the required time for color clus-

Figure 2. The impact of applying wavelet denoising on gray-scale images. From top to bottom: original image, thresholded image by Otsu's algorithm, reconstructed image using the method with wavelets [8] which is then thresholded by Otsu's algorithm

tering with less pixels. Then a combination of results is either applied or not, according to a parameter of distance and this eventual combination is partial or total in order to take into account non-connectivity of characters.

Denoising pre-processing

An important problem for thresholding methods and especially for "real-world" pictures comes from a non-uniform illumination which introduces noise. To correct this constraint, we use a wavelet decomposition as described in [8], which was a preliminary part of this entire method. The wavelet transform splits the frequency range of the gray-scale image into equal blocks and represents the spatial image for each frequency block which gives a multiscale decomposition. We use a level 8 wavelet transform using the Daubechies [1] 16 wavelet and remove low frequency subimages except the lowest one for reconstruction as shown in Figure 2.

Coarse Thresholding

The well-known Otsu method is then applied directly on the reconstructed image. This thresholding method is used for its simplicity and is sufficient for an approximate binarization at this point. Moreover this method is free-parameter and therefore general and gives all useful information on our database for a further processing. In order to keep advantage of this coarse distribution and to use color information to remove useless parts, a zonal mask is applied on the color image as shown in Figure 3 on the left. Actually in order to consider only color parts of this first threshold, a mask corresponding to useful text is applied by an AND operation on each R,G,B subimage. This task is done in order to constraint the size of the region-of-interest.

Color Clustering

In [9], color clustering is done using Graph Theoretical clustering without giving the number of clusters because the picture was not pre-processed.

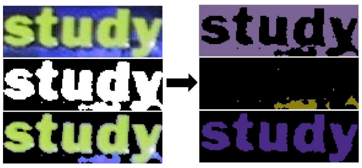

Figure 3. Right: A zonal mask applied on initial color images, left: color clustering in three subimages: background, noise and foreground

Figure 4. Two textual foreground clusters after color clustering

Actually, pre-processing with the approximate thresholding does not lose any textual information at all in our database.

We use the well-known K-means clustering with K=3. The three dominant colors are extracted based on the color map of the picture. The color map is obtained by getting all intensity values of each pixel and by removing duplicate values. Color map bins are hierarchically merged according to an Euclidean metric in the color space. These merged bins form color clusters that are iteratively updated by the K-means algorithm using the metric. Finally, each pixel in the image receives the value of the mean color vector of the cluster it has been assigned to. Three clusters are enough thanks to our pre-processing with the mask for our database. A decomposition is shown in Figure 3 on the right.

Eventual Combination

The background color is selected very easily and efficiently as being the color with the biggest rate of occurrences in the first and last lines and the first and last columns of pixels of the picture. Only two pictures left which correspond depending on the initial image to either one foreground picture and one noise picture or two foreground pictures as shown in Figure 3-right and in Figure 4.

In [9], combination is based on some texture features to remove inconvenient pictures and on a linear discriminant analysis. Here, the most probable useful picture is defined with a means of skeletonisation. Actually, as the first thresholding corresponds in an approximative way to characters, a skeletonisation is used to get the color of centers of characters as in [8].

Euclidean distance D with both mean color pixel of the cluster and mean color of the skeleton is performed. D is described in the following equation with pr1, pg1 and pb1, color values of one cluster for R,G,B channels, pr2,

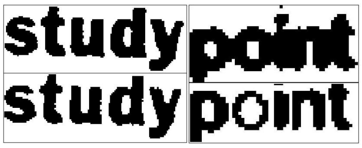

Figure 5. The first sample 'study' with well-spaced characters gives nearly the same result between our method and the Wang's one [9] on top. For the second sample 'point', our method on bottom improves the result of Wang for character segmentation

pg2 and pb2 are for the mean color value of the skeleton. The cluster with the smallest distance from the skeleton is considered as the cluster with the main textual information. Combination to do is decided according to the distance D between mean color values of the two remaining clusters.

$$D = \sqrt{(pr2 - pr1)^2 + (pg2 - pg1)^2 + (pb2 - pb1)^2}$$

If distance is inferior to 0.5, color are considered as similar and the second picture seems to be a foreground picture too. On the ICDAR database, this decision is valuable to 98.4% and no false alarms is detected. For the 1.6% remaining, some useful information is lost but the recognition is still possible as the first selected picture is the most relevant foreground one.

"Smart" Combination

Connected components on the first foreground picture are computed to get coordinates of their bounding box in order not to connect components with pixels to add in the combination. Only pixels which can be added will change the first foreground picture. On the contrary, some characters can be broken if they were broken in the first foreground picture. But, in this case, the correction will be facilitated by the fact that characters parts will be closer. Some samples are described in Figure 5 to compare final results of the Wang [9] method and our binarization technique.

Experimental results

Classical OCRs are not designed for natural scene character recognition and it is quite difficult to have pertinent results. Therefore to explain improvements our binarization method brought, we detailed results under visual judgement and by the number of connected components, which is a strong improvement for the following character segmentation.

On the ICDAR 2003 database, 21% characters are no more connected comparing to the thresholding method described in [8]. It corresponds to an improvement of 29% images and the visual judgement is drastically improved.

Concerning the Wang's method described in [9], results on binarization are quite the same but we have 6.3% characters no more connected and segmentation is improved. This improvement concerns 11.3% images of the database.

Moreover in comparison with this latter method, computation time is reduced around 40% depending on the image size. This reduction is relative to the way it can be implemented but nevertheless the time reduction is obvious.

3. Conclusion and Future Work

In this paper, we have presented a new binarization method for "real-world" camera-based pictures. Illumination and blur are corrected with a wavelet denoising. Color information is not used from the beginning in order to reduce computation time and to use the color information at a more convenient step.

Moreover a smart combination is done between clusters to get as much information as possible with a compromise with the number of connected components in order to improve character segmentation and recognition. Improvements have been done comparing to other recent techniques in binarization using color information.

A way to discriminate backgrounds between clean and noisy ones is currently under investigation to further decrease the computation time and to get smoother results in the case of clean backgrounds.

Acknowledgments

This work is part of the project Sypole and is funded by Ministere de la Region wallonne in Belgium.

References

[1] I.Daubechies: Ten lectures on wavelets, SIAM (1992)

[2] C.Garcia and X.Apostolidis: Text detection and segmentation in complex color images, Proceedings of ICASSP 2000, (2000) Vol. IV, 2326–2330

[3] J.Kittler, J.Illingworth: Threshold selection based on a simple image statistic, CVGIP, **30** (1985) 125–147

[4] Y.Liu and S.N.Srihari: Document image binarization based on texture features, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, nř5, (1997) 540–544

[5] N.Otsu: A thresholding selection method from gray-level histogram, IEEE Transactions on Systems, Man, and Cybernetics, **9** (1979) 62–66

[6] P.K.Sahoo, S.Soltani, A.K.C.Wong: A survey of thresholding technique, CVGIP, **41** (1988) 233–260

[7] M.Seeger and C.Dance: Binarising camera images for OCR, ICDAR 2001, (2001) 54–59

[8] C.Thillou and B.Gosselin: Robust thresholding based on wavelets and thinning algorithms for degraded camera images, Proceedings of ACIVS 2004, (2004)

[9] B.Wang, X-F.Li, F.Liu and F-Q.Hu: Color text image binarization based on binary texture analysis, Proceedings of ICASSP 2004, (2004) 585–588