Automatic Thresholding of Gray-level Using Multi-stage Approach

Sue Wu, Adnan Amin
School of Computer Science and Engineering
University of New South Wales
Sydney, 2052, Australia
{suewu, amin}@cse.unsw.edu.au

Abstract

A multi-stage approach is presented for thresholding document images, along with its application. The proposed method is based on two stages. Global thresholding is used in first stage to give a preliminary result. A second stage then refines the threshold value based on local spatial characteristics of the regions formed in the first stage. It automatically customizes the thresholding of regions that have specific and consistent characteristics but are different to other regions in the image. This technique works well for both simple images, in which the background and foreground are distinct and separable, and complex images containing multiple regions with different shading/textures. A typical application is postal envelope analysis. The results of evaluation show significant improvement compared to several other global and local thresholding techniques.

1. Introduction

Most document analysis algorithms are built on taking advantage of the underlying binarised image data. Information represented in two levels decreases computational load and enables the utilization of simplified analysis methods compared to 256 levels of gray-scale or color image information. Document image understanding methods require logical and semantic content preservation during thresholding.

Thresholding (Binarisation) has been a subject of an intense research interest during the last two decades [1-9,17,18] and most of the methods can be classified into two categories: global and adaptive thresholding. Global methods try to find a single threshold value for the entire image, whereas the ones in adaptive methods is calculated based on the information obtained from a certain size of neighborhood, or certain reference domain within the input image.

The two categories of thresholding methods try to find the best-fit value at once, also known as one stage thresholding. When the entire image has relative consistent characteristics, one stage thresholding gives satisfying result with appropriate method chosen, either global or adaptive. If the image is unevenly illuminated, local thresholding might outperform global method. If no priori knowledge of an image is available, global thresholding performs better. One of the limitations of local thresholding method is that it has no knowledge of important features in the image, it only determines the localized threshold for the current region that is only a small portion of the entire image. If the image has more complicated components presented, one stage thresholding is not sufficient at all. In contrast, multi stage thresholding methods take in various pieces of information to assist the classification between object of interest and background.

Several multi stage thresholding techniques has been proposed in the literature [10-13]. In [10] Chi and Yan proposed a two stages thresholding algorithm for map images by using local thresholding with different criteria. In stage one, a fixed size of window is chosen, local intensity and contrast is calculated within the window for the center pixel, then the pixel is classified against a classifying threshold value which is a constant determined experimentally. After stage one, the entire image is classified into three classes: foreground, background, and undetermined. In stage two, local intensity information is again obtained from the image for the undetermined class from previous result, then classified against new threshold value calculated by fuzzy rule implementation.

While Solihin and Leedham [11] proposed a Quadratic Integral Ratio (QIR) method which is a global two stages approach. QIR uses intensity histogram to find a threshold value for the entire image, then classify the image into three classes similar to [10]. The undetermined class is then thresholded against a threshold value calculated by an algorithm with user supplied information.

Taking the disadvantage of one stage and advantage of multi stage thresholding techniques into consideration, this

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 IEEE

paper presents a new two stages thresholding algorithm which uses spatially local information to threshold gray scale images without parameters, such as fixed window size presented in local thresholding method. The results of experimentation using several thresholding techniques including both global and local methods are presented, and compared against our new algorithm.

2. The Proposed Thresholding Algorithm

The new approach takes advantage of the multi stage thresholding technique that various information of the image needs to be considered before final threshold value is determined. The new approach consists of two stages. The first stage is to use global thresholding to identify the features of the image. Features mean region/joint segments of the image. The second stage is to perform thresholding on regions that defined those features. This approach allows the document itself to describe the area of interest. This is because the global thresholding finds the feature, then locally analysis it in second stage. For example, if you have a basket of sands with seashells in it, the first stage is to remove the sands, the seashells becomes individual images once the majority background noise is removed. Then the second stage is to pick up the individual seashells and brush the minor sands off to reveal the seashell noiseless.

The objective of our approach is to extract textual information from an image. Text is usually either darker or lighter than its background. Without loss of generality, we assume text is darker since it is the case in majority of printed text documents. Other than the above assumption, no other restriction is applied, therefore this approach is nearly parameter free.

2.1 Global Thresholding

Global thresholding is applied to find a preliminary threshold value T1. Otsu algorithm produced T1=96 for the sample image in this paper. Any global thresholding technique may be used in this step since the primary purpose is to remove some simple background that is common to the entire image. We chose Otsu's method [3] because it has been cited an effective scheme and been used in numerous document processing applications [14,15]. In Trier and Jain's OCR goal directed evaluation study [9], Otsu's method outperformed the other 3 global methods cited. Fischer's comparison of 15 global methods also confirmed that Otsu's method is preferred in document image processing [16].

For certain types of images, if the first stage fails, it is possible to implement different type of thresholding, tailored to those specific conditions. In our application, this method is used for analysis of postal letters, for which Otsu gives the best preliminary segmentation result.

The threshold value T1 is used to threshold the entire image, then connected component (CC) process is done on the resulting binary image to segment the image into smaller regions each containing one single connected component. The CC regions stand for the object of interest, the non-CC region is therefore the background.

2.2 Extraction of Sub-images

Each small region of interest formed in previous step is mapped with its gray scale values from the original image to form sub-images. For convenience, the original input image is called parent image, all sub-images extracted from the parent image are children images.

This extraction also implies a partial thresholding for the children images. That is because CC regions in binary form are the pixels having values less than global value T1 in parent image, the mapping in this step keeps CC region pixels values unchanged, but put non-CC regions pixel values to the maximum. This partial thresholding is done so that the simple background can be easily ignored in later step, and characteristic of local region of interest will appear clearer and easier to identify.

Figure 1. This child image has been mapped to original gray scale image by partial thresholding (if pixel value P(x,y)<=T1, keep original value, otherwise P(x,y)=255

2.3 Selection of Children Images Undergoing Second Stage Thresholding

In this step, all children images are tested against some criteria to see whether it needs to undergo second stage thresholding or not.

The global threshold value T1 is found based on a general assumption that there are two peaks in the intensity histogram of the parent image, one corresponding to the foreground, and the other to the background, T1 is the valley in the histogram. Otsu method uses this assumption, and works best if the pixel values in the image follow a Gaussian normal distribution. In reality, very few real life

images, including document images, strictly follow this distribution. Because the children images produced in previous step are partially thresholded images with value T1, ideally their individual histograms should have only one peak according to Otsu's method. Due to the children images' own local spatial characteristic and intensity contrast, its pixels might not follow the same distribution pattern as in the parent image. If a child image's intensity histogram has more than one peak, T1 is too high.

Only certain selections of the children images need to undergo second stage of thresholding. That is because if the object has high contrast with the image background, the first stage global thresholding is sufficient to do the binarisation. If the textual object is embedded in an uneven or low contrast background, global thresholding will have difficulty to deal with, more local information is needed for the child image, hence need to undergo second pass to refine the child image's threshold value.

2.4 Threshold Children images Independently

If a child image needs not to undergo second pass according to the test in step 3, threshold it with value T1, otherwise threshold it with the following scheme. First smooth the child image with a 5X5 mean filter, then calculate its histogram, look for the first minimum, denoted T2, in the histogram and threshold the child image on that value. Since the histogram is often very jagged, the first minimum is determined by using a 10 point moving average method. The first minimum is chosen as the refined threshold value T2 because text is darker than its background, hence forms the first peak in a child image's histogram, other subsequent peaks are formed by its uneven background pattern.

Figure 2. One of many children images undergoing second pass produced new refined thresholding value T2=35, which is more accurate than T1=96 produced by first stage.

2.5 Form Final Binary Image

All the children images resulting from the immediate previous step are placed back to their original positions, where they are extracted from the parent image, hence form the final output binary image for the original input image of this method system.

3. Experimental Result

The proposed approach for thresholding gray-scale document image has been tested on collections of postal envelope images. Over 30 unconstrained envelope samples provided by a company were used in the experiment. These images contain machine printed and hand written texts, variety of layouts, including sparse and dense textual regions, mixed fonts with different sizes and orientations, text on different shading or watermarks within one image.

3.1 Comparison With Other Thresholding Algorithms

Several other well-known methods in the literature had been used to test the same collections of images for comparison. These methods include Otsu [3], Trier and Taxt's [7], Niblack's [6], Sauvola et al.'s [17], and Wolf and Jolion's [18]. These methods are chosen because either they have been previously used to threshold document image successfully, or they were designed to extract textual information from its application. The visual evaluation of experimental results confirms that our algorithm performs better than the other cited methods. Both two-stage thresholding algorithms from reference [10, 11] are not included in the comparison because of two reasons. One is that both methods are application specific, one applies to map image, which is very different from problem domain of our approach, the other applies to hand-written images only, which requires more dedicated handling than machine print or hand writing for general purpose. Secondly, these two methods use user input to assist the decision making in either of the two stages, whereas our method is fully automated by the algorithm implemented.

3.1.1 Otsu's method. Otsu applies clustering analysis to the gray level data of the input image. The method models two clusters of Gaussian distribution of the pixels, one for background and the other for foreground regions. The optimal threshold value is found by minimizing the weighted sum of within-class variance of the two classes of pixels.

3.1.2 Trier and Taxt's method. This method is an improvement on Integrated Function method. It smoothed the input image using a mean filter, then adjust the Integrated Function method by choosing majority of pixels in its neighborhood as classifier for that individual pixel.

3.1.3 Niblack's method. Niblack suggests to calculate a threshold surface by shifting a window across the image, and use local mean and standard deviation for each center

pixel in the window. The threshold value for a pixel within fixed neighborhood is a linear function of the mean and standard deviation of the neighborhood pixels, with a constant gradient of k, which is highly tunable to separate objects well. The size of neighborhood is highly tunable, as it is chosen so that it will be small enough to preserve local details, and large enough to suppress noise.

3.1.4 Sauvola et al. Method. This method has a hybrid switch to select either text thresholding or graphics thresholding based on the properties of the image. Text thresholding switch is based on Niblack's algorithm, but differs from having a dynamic standard deviation, instead of a fixed one.

3.1.5 Wolf and Jolion's method. Wolf et al. propose to formulate the binarization decision in terms of contrast, instead of actual gray values of the pixels. Contrast in a certain neighborhood is defined in terms of a ratio, which is the difference of the neighborhood's minimum and center pixel gray values, with respect to the maximum value of standard deviations of all windows of the image. This method employs two stage approach to deal with local and global standard deviation calculation.

3.2 Result Evaluation

Figure 3 shows the original gray scale images. The information we intend to extract are:

- Name and address of the receiver
- Special request on the envelope, ie. Priority Service
- Value of the stamp

Otsu's method classifies the "Priority Service" stamp and its rectangular back-ground as a single object, the result the special request can not be recognized by any existing OCR system at all, hence loss of information.

Trier and Taxt's method causes the special request stamp fade into the background. That is because the character size of the address and the ones in the special request stamp differ too much, this thresholding can only pick up one of the two different character sizes in one pass thresholding.

Result from Niblack's method does not perform well around the region where the background is even, in other words, the algorithm tries too hard to separate pixels into two classes when there is only one class existing.

Results from Sauvola's and Wolf's methods couldn't perform well enough to binaries the "Priority Service" stamp sufficiently, also degrades the queen's head stamp.

Result from the proposed approach presented in this paper preserves the intended information very well after thresholding.

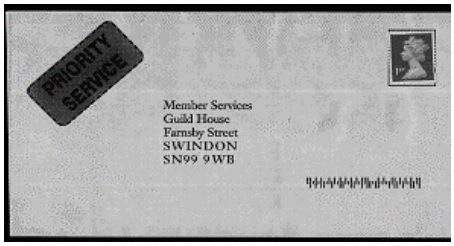

Figure 3. Original image

Figure 4. Otsu

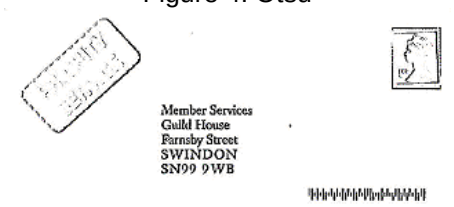

Figure 5. Trier and Taxt

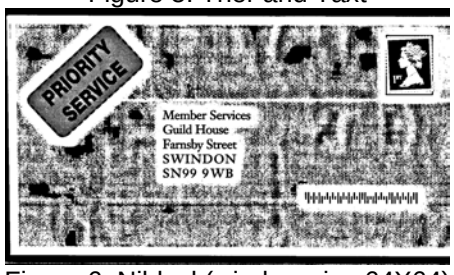

Figure 6. Niblack(window size 64X64)

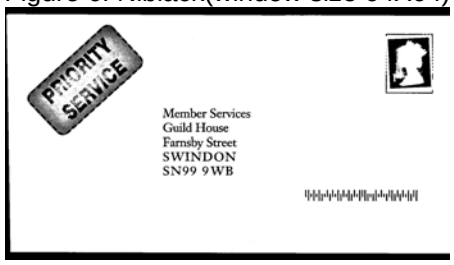

Figure 7. Salvola et al. (window size 120X120)

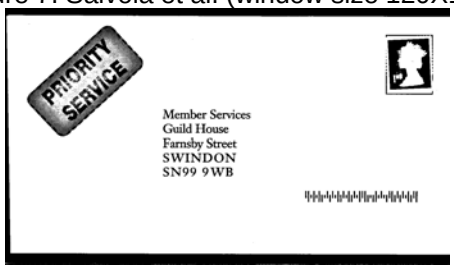

Figure 8. Wolf et al. (window size 120X120)

Figure 9. Proposal

4. Discussion

Certain document images, such as commercial envelopes processed in post offices, contain variety of background or text sizes and fonts in one single image. They require the background to be removed as much as possible and maintain as much textual information as possible for further process such as OCR. A new approach of using spatial localized two-stage thresholding strategy is presented. The technique uses a coarse to fine scheme to refine threshold values for intended objects using both global and local information obtained from the gray scale image. Experimentally, the method is compared with 5 other most referenced thresholding methods. This method appears to be a sound technique for thresholding document images, especially when the image is complicated with uneven background and different text sizes presented with different orientation etc.

The limiting factor for the method presented in this paper is to be able to segment the image into back and foreground regions. The first stage of thresholding (in our approach Otsu method) is used to run connected component analysis over the image. If the first stage failed to segment the image, then the second stage will perform no better. If the first stage failed, i.e. only produce one single image, the second stage will perform on one single image which is equivalent to the first stage. The first stage is global approach, the second stage is local approach. If the window size for local approach is too large, local approach is no difference to global approach. Therefore in our approach the first stage is to find out different local threshold window sizes for different components which form the single entire image.

The sub image formation can be viewed as dynamic window sizes chosen for local thresholding through the whole image, whilst traditional local thresholding methods use one fixed window size for the entire image.

The method is robust to poor lighting conditions because purpose of the preliminary thresholding value is to segment regions of interest, therefore the value can be determined by different thresholding methods to customized for various applications. Our sample set of test images consist of some postal envelopes scanned under dim lights, the result is satisfying.

5. References

[1] A. Abutaleb, "Automatic Thresholding of Gray-Level Pictures Using Two Dimensional Entropy", *CVGIP*, Vol. 47, 1989, pp.22-32.

[2] J. Kittler and J. Illingworth, "Minimum Error Thresholding", *Pattern Recognition*, Vol.19, pp. 41-47, 1986

[3] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram", *IEEE Trans. Systems, Man and Cybernetics*. Vol. 9, pp.62-66, 1979

[4] M. Kamel and A. Zhao, "Extraction of Binary Character/Graphics Images from Grayscale Document Images", *CVGIP*, Vol.55, No.3, pp.203-217, May 1993.

[5] W. Tsai, "Moment-Preserving Thresholding: A New Approach", *CVGIP*, Vol.29, No.3, pp.377-393, March 1985.

[6] W. Niblack, *An Introduction to Digital Image Processing*, pp. 115-116, Englewood Cliffs, N.J., Prentice Hall, 1986

[7] O.Trier and T.Taxt, "Improvement of Integrated Function Algorithm for Binarization of Document Images", *Pattern Recognition Letters*, Vol. 16, No. 3 pp.277-283, March 1995

[8] C.Glasbey, "An Analysis of Histogram-based Thresholding Algorithms", *CVGIP: Graphical Models and Image Processing*, Vol.55, No.6 pp.532-537, November, 1993

[9] O.Trier and A.Jain, "Goal Directed Evaluation of Binarization Methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 17, No.12, pp.1191-1201, 1995.

[10] Z.Chi and H.Yan, "Map Image Segmentation Based On Thresholding and Fuzzy Rules", *Electronics Letters* Vol 29, No.21, pp1841-1843, October 1993.

[11] Y.Solihin and C.Leedham, "Integral Ratio: A New Class of Thresholding Techniques for Handwriting Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol21, No.8, pp.761-768, 1999.

[12] L.O'Gorman, "Binarization and Multithresholding of Document Images Using Connectivity", *CVGIP*, Vol.56, No.6, pp.494-506, November 1994.

[13] Y. Chang, A.Fu and H.Yan, "A Hierarchical Approach in Multilevel Thresholding Based on Maximum Entropy and Bayes' Formula", *Pan-Sydney Area Workshop on Visual Information Processing (VIP2001),* Sydney, Australia, 2001.

[14] B.Yu, A.Jain and M.Mohiuddin, "Address Block Location on complex mail Pieces", *Proc. Int'l conf. Document Analysis and Recognition,* pp.897-901, Ulm, Germany, Aug. 1997

[15] A.Amin and R.Shiu, "Page Segmentation and Classification Utilizing Bottom-Up Approach", *Int'l Journal of Image and Graphics,* Vol. 1, No. 2, pp. 345-361, 2001.

[16] S.Fischer, "Digital Image Processing: Skewing and Thresholding", *Master of Science thesis*, University of New South Wales, Sydney, Australia, 2000.

[17] J. Sauvola, T. Sepplanen, S. Haapakoski, and M. Pietiklainen, "Adaptive Document Binarization", *4th International Conference on Document Analysis and Recognition*, Ulm, Ger-many, August, 1997.

[18] C. Wolf, J. Jolion, and F. Chassaing, "Text Localization, Enhancement and Binarization in Multimedia Documents", *International Conference on Pattern Recognition*, pp.1057-1040, Semptember, 2002.