MAJORITY VOTING IMAGE BINARIZATION

Alexandru PRUNCU [1]*

Cezar GHIMBAS [2]

Radu BOERU [3]

Vlad NECULAE [4]

Costin-Anton BOIANGIU [5]

ABSTRACT

This paper presents a new binarization technique for text based images. The proposed method combines several state of the art binarization algorithms through a majority voting scheme and applies a post processing effect to improve the results, more specifically, the edge map of the grayscale image is used in combination with the image resulted through the voting process in order to ensure a more accurate determination of the image's characters. Compared individually to each algorithm used, the binarization result proves to be quite promising, surpassing every other algorithm for certain images containing machine-written characters.

KEYWORDS: *image binarization, image processing, image segmentation, voting technique, edge map, Otsu, Riddler-Calvard, Niblack, Sauvola, Wolf*

1. INTRODUCTION

Image binarization refers to the process of applying thresholding to an image in order to determine which of two different color levels (usually denoted as black and white) a certain pixel will be associated with. Such an algorithm could be used in the process of finding background and foreground pixels in an image, proving itself particularly useful when dealing with text documents, usually being one of the first processing steps in any good Optical Character Recognition (OCR) software.

Classical image binarization techniques are usually separated in local and global methods. Global thresholding techniques use a single value for the whole image and are therefore faster than local ones, but only perform well in cases where there is a good separation between background and foreground. Local thresholding algorithms compute a value for

---

[1*]   corresponding author, Engineer, "Politehnica" University of Bucharest, Bucharest, Romania alexandru.pruncu@stud.acs.upb.ro

[2] Engineer, "Politehnica" University of Bucharest, Bucharest, Romania, cezar_mihai.ghimbas@cti.pub.ro

[3] Engineer, "Politehnica" University of Bucharest, Bucharest, Romania, radu_florentin.boeru@cti.pub.ro

[4] Engineer, "Politehnica" University of Bucharest, Bucharest, Romania, neculae.vlad@gmail.com

[5] Professor PhD Eng., "Politehnica" University of Bucharest, Bucharest, Romania, costin.boiangiu@cs.pub.ro

every pixel, providing better results for images with uneven lighting, but performing worse in the case of a noisy image.

On one hand, as every algorithm has different results, combining them can provide a new binarization that exploits the best parts of every individual algorithm. On the other hand, this combination can also propagate the less attractive aspects of the various algorithms, so care should be taken when deploying such a technique.

The paper is structured as follows: Section 2 presents the state of the art in terms of binarization and combining several algorithms to obtain a better result, Section 3 details the proposed algorithm that uses majority voting and a post-processing step based on the edge map of the grayscale image, Section 4 presents the results that this method produced and Section 5 describes the presented approach's conclusions.

2. RELATED WORK

An interesting approach to binarization through voting can be seen in [1]. The algorithm starts by applying a 5x5 Wiener filter [2] to the grayscale image. Using the filtered image, an odd number of binarized images are determined using various algorithms and a majority vote is applied to obtain a new binarized image. The edge map of the filtered grayscale image is determined, preserving only the edge values that are above a certain threshold. A new edge map is determined by only keeping the connected components that overlap with the voted binarized image (in a 3x3 neighborhood). The resulting shapes in the edge map are then filled in using an extension to the Run-Length Smoothing Algorithm [3] (turning successive white pixels into successive black pixels). The foreground pixels from both the resulting image and the binarized image are cumulated and a final conditional dilation is applied. This method, usually produces the best output result considering any of the input algorithms.

Voting based approaches have been previously shown to lead to promising results in other fields related to Image Processing and Computer Vision, such as OCR Systems [4], Layout Analysis [5] and Image Segmentation [6].

2.1. Local methods

Local image binarization methods compute a threshold for each "region" of the image by sliding a rectangular window over the input image.

In Niblack's approach [7], the average and dispersion of neighbors from the corresponding window are calculated for each pixel. This method can excellently recognize the foreground, but it also less resistant to noise. For Niblack, the chosen threshold is computed using the following formula:

$$T_{Niblack} = m + k * sm$$

Where m is the average, s the standard deviation and k is a constant, usually chosen (empirically) as 0.2 to balance signal with noise.

Sauvola [8] improves the previous method by using the dynamic range of standard deviation R. This works well on a light background texture and when the foreground and

background pixels are near the lower and, respectively, the upper interval range of the color range. Performance declines quickly when these values are close to each other.

$$T_{Sauvola} = m(1 - k * (1 - \frac{s}{R}))$$

Wolf's algorithm [9] solves Sauvola's drawbacks by computing a global minimum gray value, M:

$$T_{Wolf} = m(1 - k) + k * M + k * \frac{s}{R}(m - M) \tag{3}$$

Using a globally calculated value can also represent a disadvantage, as it can be influenced by noisy "regions" of the image.

2.2. Automatic thresholding

These methods start either from a random or a specific value of the threshold and, as the algorithm executes, the value is computed by using several techniques in order to obtain a better result.

A bimodal distribution is a continuous probability function with two different local maximum values. Otsu's method [10] is a globally thresholding approach which offers good results if the input image presents this type of distribution (a low minimum value between two peaks). The target image is divided into two classes (background and foreground) by choosing a value for the threshold. Then, the class mean and deviation are computed, as well as the normalized histogram. Each class has a weight associated to it, computed from the histogram's bins. Using the previous values, the method tries to minimize intra-class variance and maximize inter-class variance:

$$\sigma^2{}_b(t) = \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2$$

Where μ is the class mean and ω is the class probability. σ is repeatedly computed until the desired result is achieved.

Riddler-Calvard [11] starts from the presumption that an image is the sum of 2 distributions (background and foreground). If the 2 distributions are Gaussian (normal distributions) and their deviations are equal, then a threshold can be computed from the arithmetic mean of the distributions expected value. This new value is considered the new threshold value and the 2 new distributions are computed. The steps are repeated until the threshold value does not change (i.e. the chosen minimum error is not exceeded).

3. PROPOSED METHOD

The proposed algorithms obtain quality binarizations for input images that respect certain properties, each offering the best results for their characteristic segment. The 3 algorithms are described below.

3.1. Majority voting (unweighted)

Unweighted Majority Voting [12] consists of using multiple binarization techniques in order to generate a number of resulting images, and combining them into a single one,

setting a pixel to the value that most resulting images agree on. Each algorithm is executed and the results are compared per pixel as following: if the majority of the methods consider the pixel N_i as part of the foreground, then the final result will also be part of the foreground. The same logic applies to background pixels. An advantage of this method is that a big number of results closer to the ground truth will always result in a final result close to the ground-truth. Unfortunately, this also means that a false positive (or negative) majority will produce a wrong final image.

3.2. Majority voting (weighted)

Unlike the previous method, each algorithm's result is weighted. The value of the weight is either chosen randomly, or based on the overall results after thresholding each image in the dataset. For example, if the Wolf method's results are the closest to the ground-truth, then its vote will value more. To avoid vote manipulation, the maximum weight will not exceed the value of: *Number_of_total_methods – 2*.

The main advantage of this method is the avoidance of false positives/negatives. The biggest problem is choosing the value for the weight: random values tend to produce inconsistent results, while a specific value is highly dependent on the algorithm's implementation and the image database.

3.3. Edge based post processing

After the majority vote is done, a post-processing operation is applied. This algorithm is comprised of the following steps:

- A new background-filled image is created, in order to be used as a destination
- The edge map of the original image is determined using the Canny algorithm (after applying a Gaussian blur to ensure better results)
- Every resulting pixel in the edge map is inserted into the destination image
- Starting from the newly inserted edge pixel, the following algorithm is applied in order to find other pixels corresponding to the current shape: the edge map is being traversed to the right as long as the image resulted through voting still contains foreground pixels or until a new edge pixel is detected. During this scan, every corresponding pixel from the destination image is being written as a foreground pixel.
- The previous method is also applied vertically, starting from the same edge pixel as mentioned before.

4. RESULTS

Our test database consists of images used at the Document Image Binarization Competition (DIBCO) in 2013 [13] and 2016 [14].

Based on the value of a pixel (white or black) from the ground truth image and the resulted binarized image, they can be grouped as [15]:

- TP (true positive) - a pixel which is on in both the ground truth (GT) and the binarization result images

- FP (false positive) - a pixel which is on only in the resulted image
- FN (false negative) - a pixel which is on only in the GT image

Using this classification the following metrics can be used to compute the binarization quality:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

The complete results can be found in Table 1, whilst Table 2 presents the best, worst and edge results for several images.

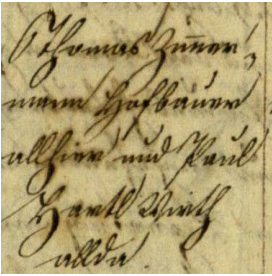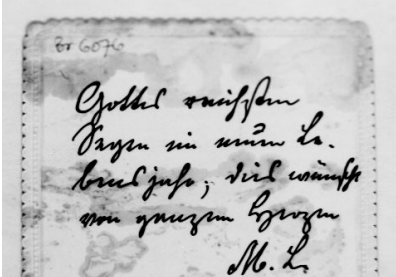Table 1. Results of the experiments. Vote is the result of a weighted majority voting process

Image	Type of degradation	Method	F-measure	PSNR	DRD
10 (DIBCO 2016)	Old, shades	Niblack	64.64	10.79	7.4
		Otsu	66.04	10.24	9.13
		Riddler-Calvard	68.24	11.1	7.15
		Sauvola	62.27	10.8	7.43
		Wolf	64.16	9.85	9.97
		Vote	**66.04**	**10.1**	**8.81**
		Edge	57.9	8.8	13.3
1 (DIBCO 2016)	Wet, stamp	Niblack	61.20	12.53	30
		Otsu	78	15.81	13.08
		Riddler-Calvard	78.17	15.8	13.13
		Sauvola	79	15.97	12.63
		Wolf	78.83	15.78	12.92
		Vote	76.16	15.56	14
		Edge	**76.74**	**15.4**	**14.41**
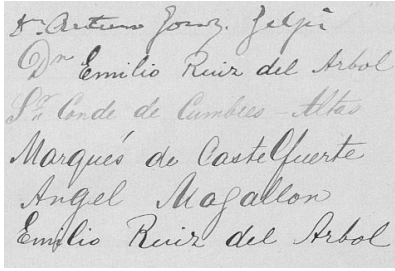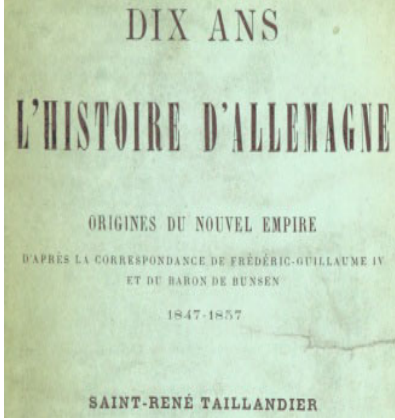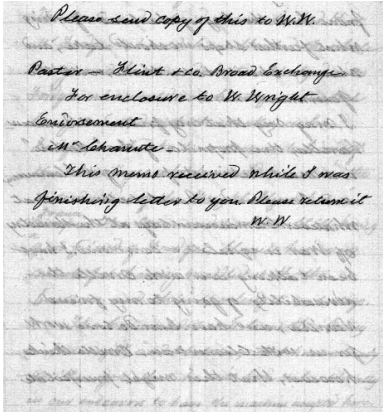
Image	Type of degradation	Method	F-measure	PSNR	DRD
7 (DIBCO 2016)	Slim faded paper	Niblack	65	12.58	9
		Otsu	57.72	12.05	10.14
		Riddler-Calvard	35	10.82	13.71
		Sauvola	39.98	11	13
		Wolf	53.34	11.75	11
		Vote	57.7	12	10.14
		Edge	**71.9**	**13**	**8.11**
PR03 (DIBCO 2013)	Noisy paper	Niblack	69.58	13.88	15.73
		Otsu	78.11	15.8	8.9
		Riddler-Calvard	65.31	14.23	12.77
		Sauvola	69.13	14.62	11.63
		Wolf	79.6	16	8.52
		Vote	77.4	15.68	9.16
		Edge	**82.27**	**16.29**	**8.42**
HW05 (DIBCO 2013)	Back page text visible	Niblack	31.39	10.75	66
		Otsu	49.96	14	28.92
		Riddler-Calvard	67.25	17.36	12
		Sauvola	62.53	16.32	15.8
		Wolf	55.37	14.93	22.44
		Vote	**50.7**	**14.29**	**27.35**
		Edge	26.86	9.8	82.64

Table.2 Best, worst and edge results for several images. Best column refers to the highest F-measure image different from our resulting ones.

Image	Edge	Best	Worst
10 (DIBCO 2016)		Riddler-C	Sauvola
1 (DIBCO 2016)		Sauvola	Niblack
Image	Edge	Best	Worst
7 (DIBCO 2016)		Niblack	Riddler-C
PR03 (DIBCO 2013)		Wolf	Riddler-C

5. CONCLUSION

Regarding the input algorithms, it's useful to say that any of them has specific advantages and drawbacks.

Riddler-Calvard tends to thin the characters, whilst produces good results when the image presents noise or back page text is visible (an example being the image "*HW05*"). However, when the text is very thin, it can misinterpret it as being part of the background

(as in image "7", where a full row of text was removed). Otsu follows the same pattern, except the letters are thicker.

Niblack's main advantage relies in the solid quality of the slim and low-contrast text, while its main disadvantage remains the excessive amount of noise introduced in the output. Both aforementioned methods, Riddler-Calvard and Otsu, taking a fill global thresholding approach instead of a full local one, are suffering from the opposite defects types.

By combining the properties of these algorithms through voting and further improving the output by filling in the gaps using edge detection, the resulting binarization images are further improved.

The results are heavily-dependent on the edge-detection algorithm, more precisely on the noise of the edge map. If the Canny edge detector result is very noisy, the post-processing step will choose a lot of irrelevant pixels and will try to fill in unimportant shapes.

The presented binarization technique, consisting of a voting based approach, usually leads to better results, when taking into account each of its composing algorithms. Adding on top of the voting binarization the post processing step further enhances the results in the majority of tests, with the exception of pictures containing old shades or where the text bleeds through from the back page. When the images to be binarized do not present these types of degradation, the Edge method should be considered, as it provides better results than the Vote one. When it is not certain whether the images contain these faults or not, the Vote method should be considered.

The problem with the latter category of degraded images lies not in the binarization method that is used, but in the way the images are interpreted. One cannot be certain if a line of text is part of the analyzed page, or part of the back page. The only clue that can help one make an educated guess is the fact that the text is written from left to right. Were the image to be reversed, the main-page text would now be considered background, and the bleed through text would become the foreground. Such exceptions should not be considered when discussing a binarization technique.

ACKNOWLEDGEMENT

REFERENCES

[1]   B. Gatos, I. Pratikakis, and S. Perantonis, "Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information," International Conference on Pattern Recognition, pp. 1–4, 2008.

[2]   Wiener N:"The interpolation, extrapolation and smoothing of stationary time series", Report of the Services 19, Research Project DIC-6037 MIT, February 1942

[3] F.M. Wahl, K.Y. Wong, R.G. Casey, "Block Segmentation and Text Extraction in Mixed Text/Image Documents", Comp. Grap. and Im. Proc., pp. 375-390, 1982

[4] Costin-Anton Boiangiu, Radu Ioanitescu, Razvan-Costin Dragomir, "Voting-Based OCR System", The Proceedings of Journal ISOM, Vol. 10 No. 2 / December 2016 (Journal of Information Systems, Operations Management), pp 470-486, ISSN 1843-4711

[5] Costin-Anton Boiangiu, Paul Boglis, Georgiana Simion, Radu Ioanitescu, "Voting-Based Layout Analysis", The Proceedings of Journal ISOM Vol. 8 No. 1 / June 2014 (Journal of Information Systems, Operations Management), pp. 39-47, ISSN 1843-4711

[6] Costin-Anton Boiangiu, Radu Ioanitescu, "Voting-Based Image Segmentation", The Proceedings of Journal ISOM Vol. 7 No. 2 / December 2013 (Journal of Information Systems, Operations Management), pp. 211-220, ISSN 1843-4711

[7] W.Niblack, An Introduction to Digital Image Processing. Prentice Hall, Englewood Cliffs, (1986).

[8] J.Sauvola, T,Seppanen, S.Haapakoski, M.Pietikainen, "Adaptive Document Binarization", 4th Int. Conf. On Document Analysis and Recognition, Ulm, Germany, pp.147-152 (1997).

[9] C. Wolf, J-M. Jolion, "Extraction and Recognition of Artificial Text in Multimedia Documents", Pattern Analysis and Applications, 6(4):309-326, (2003).

[10] N. Otsu, "A threshold selection method from grey level histogram", IEEE Trans. Syst. Man Cybern., vol. 9 no. 1, pp. 62-66 (1979).

[11] "Picture Thresholding Using an Iterative Selection Method" by T. Ridler and S. Calvard, in IEEE Transactions on Systems, Man and Cybernetics, vol. 8, no. 8, August 1978.

[12] Nabendu Chaki, Soharab Hossain Shaikh, Khalid Saeed, "Exploring Image Binarization Techniques", Springer, 2014

[13] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)", 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pp. 1471 - 476, Washington, DC, USA, 2013.

[14] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "Icfhr2016 handwritten document image binarization contest (h-dibco 2016)," in Frontiers in Handwriting Recog., International Conference on. IEEE, 2016, pp. 619–623.

[15] B. Gatos, K. Ntirogiannis and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO2009)", 10th International Conference on Document analysis and Recognition (ICDAR'09). Jul. 26-29. 2009. Barcelona. Spain. pp. 1375-1382.