WEIGHTED VOTING-BASED ROBUST IMAGE THRESHOLDING

Shahryar Rahnamayan, Hamid R. Tizhoosh, Magdy M.A. Salama

Medical Instrument Analysis and Machine Intelligence Research Group
Faculty of Engineering, University of Waterloo, Waterloo, Ontario, Canada
shahryar@pami.uwaterloo.ca, tizhoosh@uwaterloo.ca, msalama@hivolt1.uwaterloo.ca

ABSTRACT

A new robust image thresholding technique is introduced in this paper. Comprehensive experiments show that a single thresholding method can not be successful for all kind of images. The proposed approach uses fusion of some well-known thresholding methods by applying weighted voting at the decision level. The main objective is improving robustness of thresholding approach by participating several methods. Although, the proposed approach can not guaranty the best result for all kind of images but it shows higher performance and consistent/smoother behavior in overall. The performance of the new approach and nine well-established thresholding methods are compared by applying to an image set with high image diversity. The comparison results show that the proposed approach outperforms other nine well-established thresholding approaches. The proposed approach has been explained in details and experimental results are provided.

KEY WORDS: Thresholding, Segmentation, Voting, Misclassification Error, Kittler, Fusion

1. INTRODUCTION

Thresholding or binarization is a challenging task in image processing field. In most image processing chains, the role of this crucial task can be observed. Even in cases which other tasks use thresholding result as an input, the final performance of that processing chain is directly dependent on the result of thresholhing [1], such as character recognition in text documents. Although, many thresholding techniques have been proposed but in fact many of them are application or domain oriented solutions [2]. Robustness is a desirable factor in all science and engineering fields. For our case, it means the level of adaptation capability of the approach to the widest range of images. After comprehensive evaluation of more than 40 image thresholding techniques, Sezgin and Sankur [3] concluded *"It was observed that any single algorithm could not be successful for all image types, even in a single application domain."* According to the reported ranking in their work, the Kittler [4] is the best performing method (ranked first) and the Brink [5] is the worst one (ranked $40th$). Fig.1 shows an example to confirm their concluding sentences: The worst approach, Brink, shows a better result than the first ranked approach, Kittler. In this paper, we would like to achieve higher robustness by fusing more thresholding methods at the decision level by applying weighted voting.

The structure of the paper is as follows: In Sec.2 we talk about the proposed approach. In Sec.3 the performance criteria to evaluate thresholding results is introduced. Results, conclusion remarks, and directions of future work are given in Sec. 4, 5, and 6, respectively.

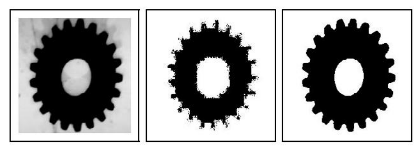

Fig. 1. Left to right: Input image, thresholded image by Kittler, and thresholded image by Brink.

2. PROPOSED APPROACH

The proposed approach is straightforward and has four main steps as follows:

(1) Methods Selection - We need to choose some successful thresholding methods to participate in our election. In order to achieve a higher performance in final thersholding results, some well-established methods should be selected according to their ranking. This paper utilizes the results of the previous mentioned comprehensive study [3] for selecting these methods and also to assign weights for voting (the next step of the approach). In that work, Sezgin and Sankur compared 40 thresholding methods quantitatively and finally ranked them according to the average of five criteria of shape segmentation goodness. We select nine top methods from their ranking. The reason to select "nine" methods will be discussed in step 4. These methods and their overall error score are given in Table 1. The Average Error Score (AES) is the mean of five errors, namely, misclassification error, region

Table 1. Nine selected top thresholding methods and their average error scores (AES). The ranking and also overall error scores are borrowed from Ref. [3].

Rank	*Method*	*AES*
1	Kittler [4]	0.256
2	Kapur [6]	0.261
3	Sahoo [7]	0.269
4	Yen [8]	0.289
5	Lioyd [9]	0.292
6	Otsu [10]	0.318
7	Yanni [11]	0.328
8	Yanowitz [12]	0.339
9	Hertz [13]	0.351

nonuniformity, relative foreground area error, and shape distortion penalty via Hausdorff distance. As appeared in Table 1, the nine selected methods are Kittler (ranked first), Kapur, Sahoo, Yen, Lioyd, Otsu, Yanni, Yanowitz, and finally Hertz method.

(2) Thresholding Input Image by Individual Methods - In this step, input gray-level image is tresholded by each selected method individually. Following names are assigned to resulted images:

Kittler: I_1, Kapur: I_2, Sahoo: I_3, Yen: I_4, Lioyd: I_5, Otsu: I_6, Yanni: I_7, Yanowitz: I_8, and Hertz: I_9.

(3) Assigning Weight or Number of Votes - In order to apply weighted voting instead of majority voting, assigning weights/number of votes for each individual method is required. As a reasonable way, we assign number of votes for each method according to its performance. In other words, methods with higher performance will receive higher vote numbers/weights. As shown in Table 1, methods with smaller average error score have been received higher ranks. We use average error score as a measure to assign our weights. Fig.2 shows average error score for selected nine thresholding methods. The curve is much almost linear. This feature has been used to calculate weights. Also, other approaches can be applied to calculate weights but because of mentioned linearity the following straightforward way is used . In the following equation, the slope of AES curve has been used to calculate weights for each method:

$$W_j = \lceil W_k + \frac{(AES_k - AES_j)}{\frac{(AES_k - AES_1)}{(k-1)}} \rceil, \quad (1)$$

where W_k and AES_k are the weight and the average error score of the last method, the $9th$ in our ranking. We assign two votes for Hertz method to start, $W_k = 2$. In the same way, W_j and AES_j are the weight and the average error score of jth method. Finally, AES_1 is the average error score of the first ranked method, which is the Kittler with $AES = 0.256$. By applying Eq.1, following weights have been achieved:

$W_{Kittler} = 10$, $W_{Kapur} = 10$, $W_{Sahoo} = 9$, $W_{Yen} = 8$, $W_{Lioyd} = 7$, $W_{Otsu} = 5$, $W_{Yanni} = 4$, $W_{Yanowitz} = 4$, and $W_{Hertz} = 2$.

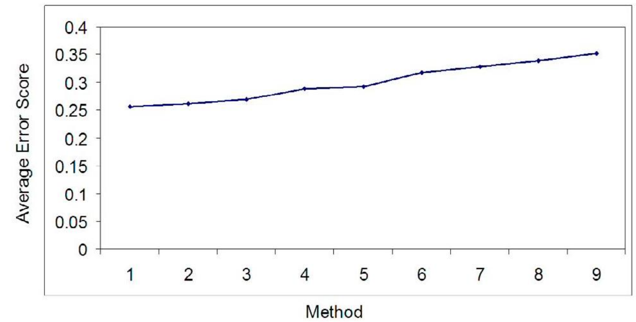

Fig. 2. Average error score (AES) for selected nine thresholding methods.

Now, we have all weights and ready to apply weighted voting approach to obtain final thresholded image.

(4) Weighted Voting - In order to decide about being zero or one for each pixel (binarization), I_{ij}, in the final threshold image, I, weighted median [15] has been used to implement weighted voting idea. For making this decision all corresponding pixels from result of individual tresholding methods participate ($I_{1_{ij}}$, ..., $I_{9_{ij}}$). Following equation shows pixel-by-pixel decision of weighted median to generate final threshold image:

$$I_{ij} = median\{W_1 \times I_{1_{ij}}, W_2 \times I_{2_{ij}}, ..., W_9 \times I_{9_{ij}}\}, \quad (2)$$

where W_k is the corresponding weight of method k. The symbol $\times$ denotes the replication operator and $W_k \times I_{k_{ij}}$ means W_k times repetition of bit $I_{k_{ij}}$ in the decision set. Note, sum of votes/weights, $\sum W_k = 59$, is odd to provide a central value.

3. PERFORMANCE CRITERION

In order to evaluate the performance of different methods, an objective metric is required. To compare two binary images, Misclassification Error (ME) [14] can be a reasonable measure. ME calculates percentage of foreground pixels which assigned wrongly to background and vice versa. By utilizing this error measure, similarity index, η, can be defined as follows:

$$\eta = \{1 - \frac{K \times (|B_O \cap F_T| + |F_O \cap B_T|)}{|B_O| + |F_O|}\} \times 100\%, \quad (3)$$

Table 2. Average similarity index ($\bar{\eta}$) and standard deviation (σ) of each method over 15 test images.

Method	$\bar{\eta}$ (%)	σ (%)
Kittler	85.31	13.17
Kapur	77.17	30.87
Sahoo	83.90	29.41
Yen	83.76	30.40
Lioyd	77.90	34.50
Otsu	78.26	23.38
Yanni	82.18	12.65
Yanowitz	76.70	24.06
Hertz	69.02	37.25
New Approach	**90.31**	**9.52**

where B_O, F_O, B_T, and F_T are the background and foreground pixels of the ground-truth image and the background and foreground pixels of the test image, respectively. $|.|$ denotes the cardinality of the set. K is a dissimilarity magnifying factor. The number of unmatched pixels is multiplied with K, $K = 5$ here, to magnify the dissimilarity between the resulting image and the ground-truth image. The level of magnification was determined empirically to match the visual expectation more realistically (a magnification factor of 1 would result in very high similarity even if many pixels are not matched).

4. EXPERIMENTAL RESULTS

In order to evaluate robustness of each method, an image set with higher image diversity was prepared. We selected 15 completely different and hard to threshold images. By applying each method to our test set, average similarity index ($\bar{\eta}$) (see Sec.3) and standard deviation are calculated and summarized in Table 2. The proposed approach shows the highest average similarity index of 90.31% with the lowest deviation of 9.52%. Fig.3 shows this behavior graphicly. As seen, for most of images, the proposed approach is located on the top of other methods and its curve is the flattest one. By this way, proposed weighted voting outperforms other nine methods. Thresholding results of some sample images are given in Fig.4. As seen, by visual (subjective) assessment, for first image (stones image), proposed approach shows better thresholding result than results of Kittler, Kapur, Yen, Lioyd, Otsu, Yanni, Yanowitz, and Hertz, compared with corresponding ground-truth image. For second image, it performs better than Kittler, Lioyd, Otsu, Yanni, Yanowitz, and Hertz. For third one, it shows better result than Kapur, Sahoo, and Yen. For fourth image, that is better than Lioyd, Otsu, Yanni, Yanowitz, and Hertz.

5. CONCLUSION REMARKS

In this paper, a weighted voting-based bilevel image thresholding approach has been proposed. Main objective was the developing a robust image thresholding approach. Robustness here means capability of delivering satisfactory result for a higher variety/diversity of images (images from different categories). Although, the proposed approach can not guaranty the best result for all cases but it presents higher performance and consistent/smoother behavior. The proposed approach can be assumed as a sort of fusion method which is applied at the decision level. Note that this approach is time consuming because it uses results of other methods to make a final decision. This approach can be used in applications which depend more on higher performance rather than prompt solution.

6. FUTURE WORKS

Quantitative performance comparison of the proposed approach with a comprehensive set of thresholding methods and dynamic adjustment of voting weights according to image characteristics are the directions of our future work.

7. REFERENCES

[1] S. Rahnamayan, H.R. Tizhoosh, M.M.A Salama, "Automated snake initialization for the segmentation of the prostate in ultrasound images," Int. Conf. on Image Analysis and Recognition (ICIAR-05), Springer Lecture Notes in Computer Science Series, Toronto, Canada, pp. 930-937, 2005.

[2] S. Rahnamayan, H.R. Tizhoosh, M.M.A Salama, "Image thresholding using differential evolution," accepted in International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV-06), Las Vegas, USA, 2006.

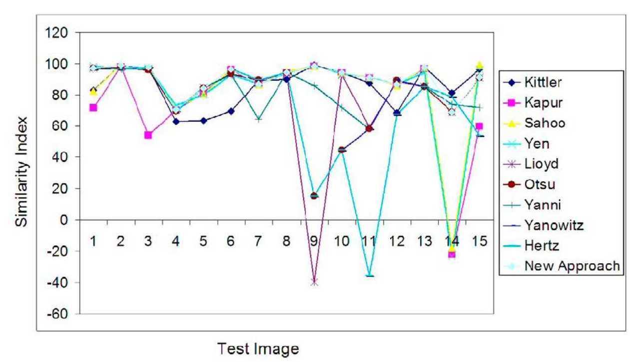

Fig. 3. Similarity index of nine selected methods and proposed method over 15 test images.

[3] Mehmet Sezgin and Bulent Sankur, "Survey over image thresholding techniques and quantative performance evaluation," Journal of Electronic Imaging 13(1), pp. 146-165, Jan. 2004.

[4] J. Kittler and J. lllingworth, "Minimum Error Thresholding," Pattern Recognition, vol. 19, pp. 41-47, 1986.

[5] A. D. Brink and N. E. Pendock, "Minimum cross entropy threshold selection," Pattern Recogn. 29, pp. 179-188, 1996.

[6] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," Graph. Models Image Process. 29, pp. 273-285, 1985.

[7] P. Sahoo, C. Wilkins, and J. Yeager, "Threshold selection using Renyis entropy," Pattern Recogn. 30, pp. 71-84, 1997.

[8] J. C. Yen, F. J. Chang, and S. Chang,"A new criterion for automatic multilevel thresholding," IEEE Trans. Image Process. IP-4, pp. 370-378, 1995.

[9] A. Rosenfeld and P. De la Torre, "Histogram concavity analysis as an aid in threshold selection," IEEE Trans. Syst. Man Cybern. SMC-13, pp. 231-235, 1983.

[10] N. Otsu, "A threshold selection method from gray level histograms," IEEE Trans. Syst. Man Cybern. SMC-9, pp. 62-66, 1979.

[11] M. K. Yanni and E. Horne, "A new approach to dynamic thresholding," 9th European Conf. Sig. Process. 1, pp. 34-44, 1994.

[12] S. D. Yanowitz and A. M. Bruckstein,"A new method for image segmentation," Comput. Graph. Image Process. 46, pp. 82-95, 1989.

[13] L. Hertz and R. W. Schafer, "Multilevel thresholding using edge matching," Comput. Vis. Graph. Image Process. 44, pp. 279-295, 1988.

[14] W. A. Yasnoff, J. K. Mui, and W. Bacus, "Error measures for scene segmentation," Pattern Recognition 9, pp. 217-231, 1977.

[15] M. Fischer, J. L. Paredes, G. R. Arce, "Image sharpening using permutation weighted median," Conf. Proc. of EUROSIP-2000, Vol. 1, pp. 299-303, Tampere, Finland, Sep. 2000.

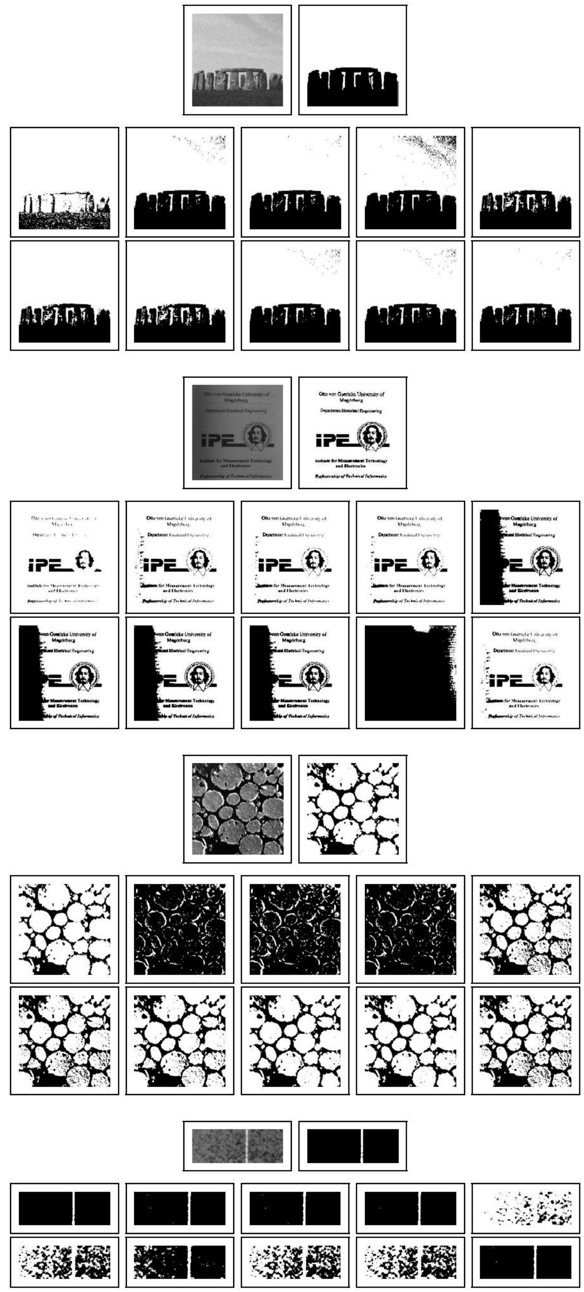

Fig. 4. Thresholding results of some sample images. For each image group, left to right: First row: {Input image and ground-truth image}, Second row: {Results of Kittler, Kapur, Sahoo, Yen, and Lioyd}, Third row: {Results of Otsu, Yanni, Yanowitz, Hertz, and new approach}.