A Weighting Scheme for Improving Otsu Method for Threshold Selection

Hui-Fuang Ng[1], Cheng-Wai Kheng[1], and Jim-Min Lin[2]

[1] Department of Computer Science, Universiti Tunku Abdul Rahman,

Kampar 31900, Perak, Malaysia

{nghf, khengcw}@utar.edu.my

[2] Department of Information Engineering and Computer Science, Feng Chia University,

Taichung 407, Taiwan, ROC

jimmy@fcu.edu.tw

Received 14 September 2015; Revised 16 September 2015 ; Accepted 16 September 2015

Abstract. Thresholding is an important technique for image segmentation that extracts a target from its background on the basis of the distribution of gray-levels. One of the popular automatic threshold selection methods, the Otsu method, provides satisfactory results for thresholding images with obvious bimodal gray-level distribution. However, Otsu method fails if the foreground and background pixels have significantly different variances, or if the histogram is unimodal or close to unimodal. Valley-emphasis method partially resolves such problems by weighting the objective function of the Otsu method using the probability information at the valley-point in the histogram. In this study, a new weighting scheme is proposed for improving Otsu threshold selection which incorporates a measure of valley deepness with probability of occurrence at the threshold location to enhance the weighting effect. Experimental results indicate that the proposed method greatly improves the performance of the Otsu method and is highly competitive with other widely used thresholding methods.

Keywords: image segmentation, image thresholding, Otsu method, valley-emphasis method

1 Introduction

Image thresholding is a commonly used technique for image segmentation because of its simplicity and computational efficiency [5]. The basic idea of image thresholding is to select an ideal threshold value based on gray-level distribution for separating objects of interest in an image from the background, assuming that the gray-levels are substantially different between the foreground objects and the background. Because of its wide applicability to other areas of image processing and applications such as automated visual inspection [11] and medical image segmentation [18], many automatic thresholding algorithms have been proposed in the literatures. In-depth survey and evaluation of thresholding methods are given by Sahoo et al. [16], Lee et al. [9], Glasbey [3], and more recently, by Sezgin and Sankur [17].

Among the image thresholding techniques, Otsu method [13] is one of the better threshold selection methods for general real world images with respect to uniformity and shape measures [16]. Due to its popularity, this method is adopted in some free and commercial software such as GIMP (www.gimp.org) and MATLAB (MathWorks, Inc.) as the default threshold selection method. Otsu method assumes that the gray-level of the object and the background in an image distribution is Gaussian distribution with equal variances and selects threshold values that maximize the between-class variances of the histogram. Therefore, Otsu method is optimal for thresholding a histogram with distinct bimodal or multimodal distribution. However, in general, real world images rarely possess such characteristics. It has been shown that Otsu method biases toward the component with larger within-class variance [20]. In other words, Otsu method fails if the foreground and background pixels have significantly different variances, or if the histogram is unimodal or close to unimodal. Several modified versions of Otsu method have been proposed, such as valley-emphasis method [12], neighborhood valley-emphasis method [4], minimum variance method [6], variance and intensity contrast method [15], Rayleigh distribution-based method [19], median-based method [21,23], and adaptable threshold detector [14].

The valley-emphasis method [12] uses similar objective function as the Otsu method for finding optimal threshold value but gives more weights to locations in the histogram that have small probability of occurrence (valley-points). The weight is defined as one minus the probability value at the threshold location. Applying such weight in the objective function makes the threshold closer to the actual valley of the histogram. That is, valley-emphasis method maximizes between-class variances but at the same time favors threshold value near the valley-point of bimodal distribution, or at the bottom rim of unimodal distribution. As a result, this method is able to select optimal threshold value for both bimodal and unimodal distributions and has been shown to be effective in industrial defect detection applications [12]. Recently, Fan and Lei [4] pointed out that since valley-emphasis method uses only the probability value at a point to compute the weight, the weighting effect might not be strong enough for some cases to correct for the deviation of the Otsu threshold value from the true valley-point of the histogram, thus producing incorrect thresholding results. They suggested that including the neighboring information around a threshold point in the weight calculation could improve the weighting effect. However, it is unclear how to determine the appropriate size for the neighborhood to achieve optimal thresholding results.

In this paper, we extend the valley-emphasis method and propose a new weighting scheme to enhance the weighting effect for valley-points in the objective function. The new weighting scheme incorporates a measure of valley deepness with probability of occurrence in the weight calculation. As shown in the experimental results, the proposed method greatly improves the accuracy and stability of the original Otsu method.

2 Image Thresholding Methods

In this section, we briefly review the Otsu method and the valley-emphasis method for selecting optimal image threshold, and present the new weighting scheme for enhancing the weighting effect for valley-points.

2.1 Otsu Method

An image can be represented by a 2D gray-level intensity function $f(x, y)$. The value of $f(x, y)$ is the gray-level, ranging from 0 to L-1, where L is the number of distinct gray-levels. Let the number of pixels with gray-level i be n_i, and n be the total number of pixels in a given image, the probability of occurrence of gray-level i is defined as:

$$p(i) = \frac{n_i}{n} \tag{1}$$

The average gray-level of the entire image is computed as:

$$\mu_T = \sum_{i=0}^{L-1} ip(i) \tag{2}$$

In the case of single thresholding, the pixels of an image are divided into two classes $C_1 = \{0, 1, \ldots, t\}$ and $C_2 = \{t+1, t+2, \ldots, L-1\}$, where t is the threshold value. C_1 and C_2 are normally corresponding to the foreground (objects of interest) and the background. The probabilities of the two classes are:

$$\omega_1(t) = \sum_{i=0}^{t} p(i) \tag{3}$$

$$\omega_2(t) = \sum_{i=t+1}^{L-1} p(i) \tag{4}$$

The mean gray-level values of the two classes can be computed as:

$$\mu_1(t) = \sum_{i=0}^{t} ip(i) / \omega_1(t) \tag{5}$$

$$\mu_2(t) = \sum_{i=t+1}^{L-1} ip(i) / \omega_2(t) \tag{6}$$

Using discriminant analysis, Otsu [13] showed that the optimal threshold t^* can be determined by maximizing the between-class variance σ_B; that is:

$$\begin{aligned} t^* &= Arg \underset{0 \le t < L}{Max} \{\sigma_B) \\ &= Arg \underset{0 \le t < L}{Max} \{\omega_1(t)(\mu_1(t) - \mu_T)^2 + \omega_2(t)(\mu_2(t) - \mu_T)^2\} \end{aligned} \tag{7}$$

Eq. (7) could be further simplified as (Liao et al. [10]):

$$t^* = Arg \underset{0 \le t < L}{Max} \{\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t)\} \tag{8}$$

From Eq. (8) we can see that Otsu method is simple and easy to realize thus makes it one of the most commonly used threshold methods in engineer practices with satisfactory results. However, Otsu method works well for thresholding a histogram with bimodal with equal variances but fails if the histogram is unimodal or close to unimodal.

2.2 Valley-Emphasis Method

In the case of single thresholding, the idea threshold value should lie at the valley of the bimodal distribution, as shown in Fig. 1. Thus the probability at the threshold value has to be small.

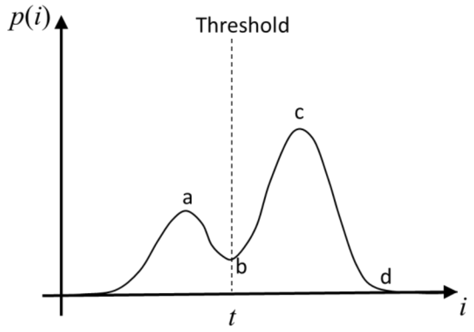


Fig. 1. Optimal threshold selection in gray-level histogram

Based on this observation, Ng [12] proposed the valley-emphasis method to select a threshold value (t) with a small probability of occurrence ($p(t)$) which also maximizes the between-class variance. They introduced a weighting term which is defined as:

$$W(t) = 1 - p(t) \tag{9}$$

The optimal threshold is chosen by maximizing the weighted objective function as:

$$t^* = Arg \underset{0 \le t < L}{Max}\{W(t)(\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t))\} \tag{10}$$

The first term in Eq. (10) is the valley weight and the second term is the between-class variance of the gray-level distribution. The smaller the $p(t)$ value, the larger the weight. This weight makes the resulting threshold closer to the valley-point of the gray-level distribution. It was shown in Fan and Lei [4] that the weighting term in Eq. (10) might not be significant enough for some cases to correct for the deviation of the Otsu threshold value from the true valley-point of the histogram such that it fails to determine the correct threshold value. This is mainly due to the fact that using only the probability of occurrence information for weight calculation might not be sufficient. For instance, point *d* in Fig. 1 has smaller probability value than the value of the true valley-point *b*. Which means point *d* will receive larger weight (1-$p(t)$) than point *b* even though it is not a valley-point. Therefore, additional information is needed for identifying and assigning appropriate weights to the truth valley-points in the histogram. To this end, in the following we propose a new weighting scheme which includes a measure of the valley deepness in addition to probability of occurrence in the weight calculation.

2.3 Proposed Method

A valley-point should have two immediate peaks around it, one to its left and another to its right, as shown in Fig. 1. Davies [2] proposed several measures for computing valley deepness in gray-level distributions. Here, we adopt Davies' ideas and devise a similar measure that is suitable for our weighting scheme. For any potential valley-point b in a histogram (Fig. 1), for a point a on the left of b ($a<b$) and a point c on the right of b ($b<c$), and take account of the corresponding heights $p(a)$, $p(b)$, $p(c)$ in the histogram, we define the left valley deepness lD and right valley deepness rD at point b as:

$$lD(b) = \max_{0 \le a < b}\{s(p(a) - p(b))\} \tag{11}$$

$$rD(b) = \max_{b < c < L}\{s(p(c) - p(b))\} \tag{12}$$

Where $s()$ is the sign function such that $s(u) = u$ if $u>0$ and $s(u) = 0$ if $u \leqq 0$. The sign function is used to prevent negative responses. The overall valley deepness D at point b is defined as:

$$D(b) = \begin{cases} [lD(b) + rD(b)]/2 & \text{if} \quad lD(b) > 0 \text{ and } rD(b) > 0 \\ 0 & \end{cases} \tag{13}$$

Essentially, the valley deepness measure in Eq. (13) represents the average distance from a valley-point to the left and right highest peaks. Using this measure, point b in Fig. 1 should have a high score since it resides between two large peaks, and point d will have zero score because there is no peak on its right thus it is not considered as a valley-point. We can see that the valley deepness measure provides accurate information for identifying valley-points in a histogram and also a means for comparing between valley-points. It should be noted that valley deepness measurement involves computing the differences between points thus it is susceptible to local noise in the distribution. As a result, a 1D Gaussian filter is used to pre-smooth the histograms before all valley deepness calculations.

In this study, we propose a weighting scheme which incorporates valley deepness measure with probability of occurrence for threshold selection. For each candidate threshold location t, the new weight is defined as:

$$W(t) = (1 - p(t)) + D(t) \tag{14}$$

The first term in Eq. (14) is the original weight used in the valley-emphasis method. The second term weights the valley deepness at the threshold location. The first weight forces the threshold to stay as low in the gray-level distribution as possible, and the second weight makes certain that the selected threshold is the global valley. The two terms are necessary and complement each other since two valley-points might have similar valley deepness but with different values of probability of occurrence. When this happens, the one with smaller probability will receive higher weight. During threshold selection, simply replace the weight term in Eq. (10) with the new weight defined in Eq. (14) and search for a threshold t that maximizes the weighted objective function.

One advantage of the valley-emphasis method is that it does not add any free parameter or computational overhead to the original Otsu method. Similarly, the proposed method does not introduce any new parameter to the objective function. In addition, valley deepness $D(t)$ can be computed efficiently in O(N) time using a two-passes strategy [2], where N is the number of gray-levels. Therefore, the computational overhead of proposed method is small.

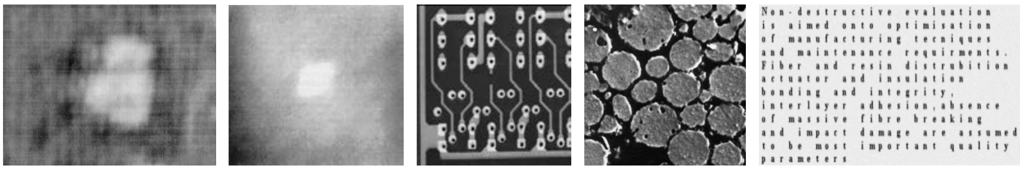
3 Experiments

In the experiments, the performance of the proposed image thresholding approach is evaluated using real-world images with ground truth. The results of the proposed method are compared with the results of Otsu method, valley-emphasis method, several other modified versions of Otsu method, as well as those obtained from widely used thresholding methods and some new methods.

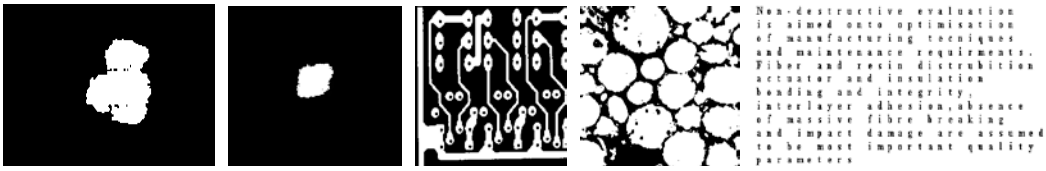
3.1 Test Images

The test dataset consists of a wide variety of image types including printed circuit board (PCB), eddy current, thermal, ultrasonic, microscope cell and material, textile, and document images, with their corresponding ground truth. The dataset is provided by Sezgin and Sankur [17] for image segmentation research and it is publicly available (www.mehmetsezgin.net). The dataset had been used in recent image thresholding studies [1,21]. Fig. 2 shows samples of the test images and their corresponding ground truth.

The original dataset consists of 25 images. Beauchemin [1] selected 22 images for their experiments, correspond to number 1–16, and 18–23 in the original dataset. The three images removed have properties similar to some of the remaining images. The reason for doing so is to avoid biases in the performance analysis toward methods that perform better (or worse) on one particular type of images. For comparison purposes, in our experiments, we use the same 22 images as in Beauchemin [1].



(a) Sample test images



(b) Ground truth

Fig. 2. Sample test images and their corresponding ground truth

3.2 Thresholding Performance Evaluation

Quality of thresholding result is quantitatively evaluated by misclassification error (*ME*) measure, which regards image segmentation as a pixel classification process. *ME* is defined as [22]:

$$ME = 1 - \frac{|B_0 \cap B_T| + |F_0 \cap F_T|}{|B_0| + |F_0|} \tag{15}$$

Where B_0 and F_0 denote the background and foreground of the original image, B_T and F_T denote the background and foreground of the test image, and |.| is the cardinality of the set. *ME* reflects the percentage of background pixels wrongly assigned to foreground, and conversely, foreground pixels wrongly assigned to background. The value of *ME* varies between 0 for a perfectly classified image and 1 for a totally erroneously classified one. A lower value of *ME* means better quality of corresponding thresholded image.

3.3 Experimental Results

Fig. 3 shows the thresholding results of one test image (image #20) of the dataset using the original Otsu method, the valley-emphasis method, and the proposed method. We can see that the thresholded image of the proposed method is very similar to the ground truth image. The result is better than that obtained by Otsu method and the valley-emphasis method. The *ME* values of the Otsu method and the valley-emphasis method are 0.070 and 0.054 respectively, while the *ME* value of the proposed method is 0.015, which is much smaller than that of previous two methods.

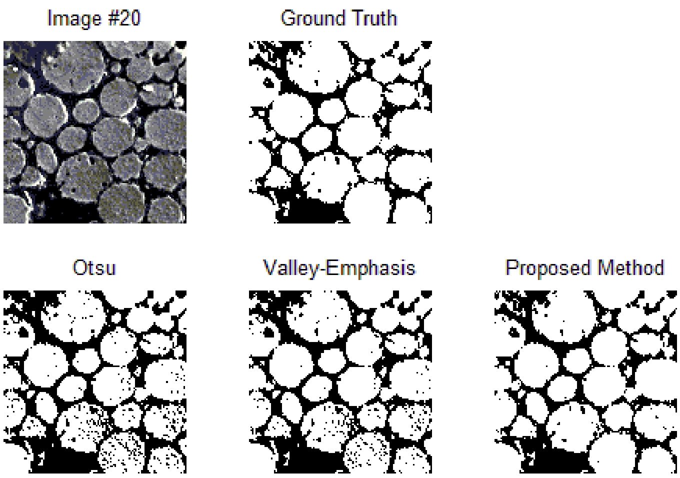


Fig. 3. Thresholding results of image #20

Fig. 4 shows the histogram of image #20. The vertical lines indicate the threshold values obtained by the three threshold methods. The histogram shows two clear modes with different variances. The threshold value selected by Otsu method is 79, which is not at the valley (around 60) of the distribution. The threshold value selected by valley-emphasis method is 75, which is slightly closer to the valley but still deviate from the valley-point. This confirms that the weighting effect of the valley-emphasis method is not large enough to rectify the error of Otsu threshold when the variances of the modes are very different. The proposed method reports a threshold value of 59, which is very close to the valley-point of the histogram. Fig. 5 shows the weight functions produced by the valley-emphasis method and the proposed method. We can see that the curve of the valley-emphasis weight is rather flat and has pedestals at either end of the distribution. And since the probability at the valley-point is not zero, the valley-emphasis weight does not peak at the valley-point. This greatly weakens the weighting effect. On the other hand, the proposed weighting function shows strong peaks at the valley points (60 and 215). The deeper the valley is, the larger the weight. Therefore, the proposed weighting scheme is able to produce more accurate thresholding results.

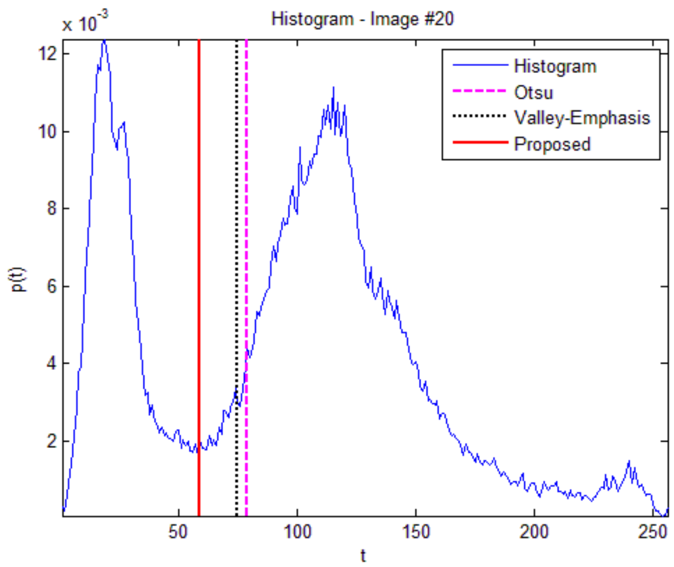


Fig. 4. Histogram of image #20

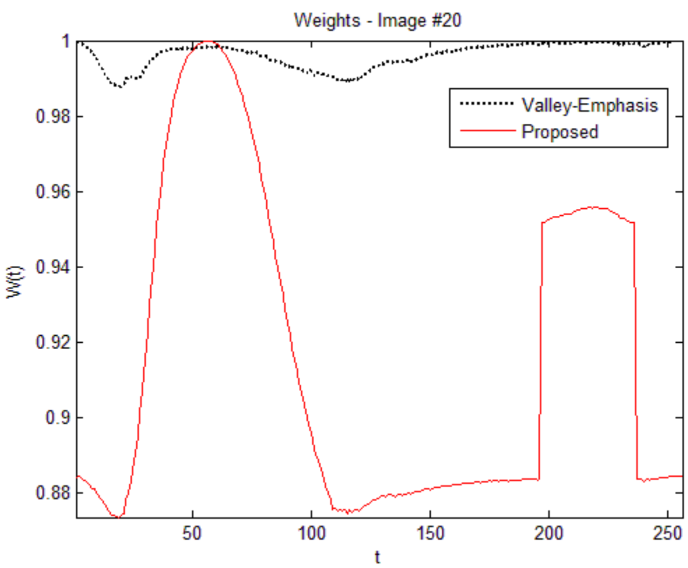


Fig. 5. Weight functions for image #20

Fig. 6 shows the thresholding results of another test image (image #5). For this image, both of the Otsu method and valley-emphasis method fail to threshold the image correctly. The *ME* values are 0.630 and 0.644 for the Otsu method and the valley-emphasis method, respectively. The proposed method, on the other hand, produces correct thresholding results with a *ME* value of 0.005.

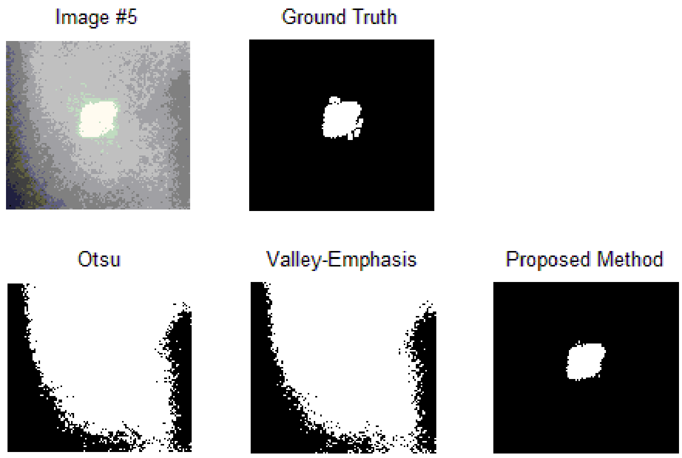


Fig. 6. Thresholding results of image #5

From the histogram of the image #5 in Fig. 7, we can see that the size and variance of the foreground are significantly smaller than that of the background. Otsu method normally fails in such cases [20], and the valley-emphasis weight is unable to rectify the error. As shown in Fig. 8, the proposed weighting function peaks at the valley-point (around 225) and provides adequate weights to shift the threshold closer to the true valley-point.

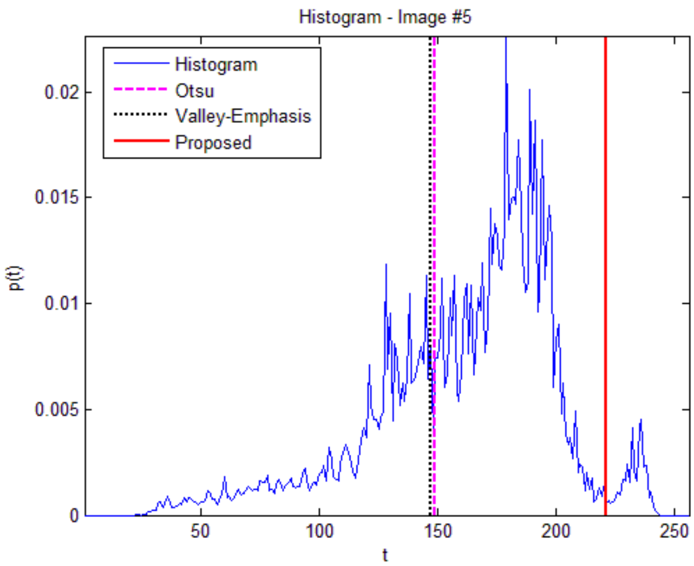


Fig. 7. Histogram of image #5

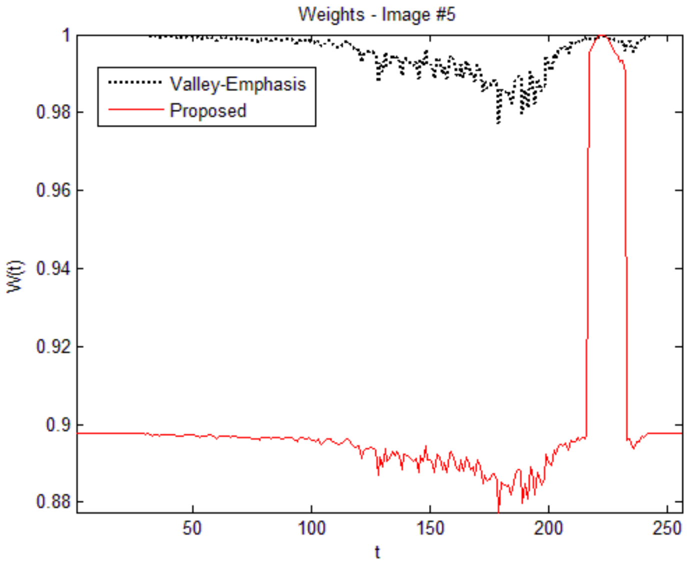


Fig. 8. Weight functions for image #5

The average misclassification errors (*ME*) of the 22 test images of various thresholding techniques are shown in Table 1. The thresholding methods included in the experiments are: 1) Otsu method [13]; 2) Valley-emphasis method [12]; 3) Neighborhood valley-emphasis method [4]; 4) Minimum variance method [6]; 5) Rayleigh distribution-based method [19]; 6) Median-based method [21]; 7) Global valley method [2]; 8) Minimum error threshold (MET) method [8]; 9) Kapur method [7]; 10) Semivariance method [1]; and 11) Proposed method. Methods 2-6 and the proposed method are considered as extended versions of the Otsu method. For method 3, the neighborhood size is set to 11 as suggested by the authors. Method 7 directly uses the valley location as threshold. Methods 8 and 9 are two widely used thresholding methods and they are ranked best methods in a comprehensive survey of image thresholding conducted by Sezgin and Sankur [17]. Method 10 is a recently proposed method based on semivariance similarity between original image and thresholded image. The method reports good threholding results on the same image dataset used in this study thus it is included here for comparison.

Table 1. Performance of thresholding methods based-on misclassification error

Method	Misclassification Error (ME)	
	Mean	**STDEV**
Otsu	0.174	0.209
Valley-emphasis	0.078	0.149
Neighborhood valley-emphasis	0.139	0.277
Minimum variance	0.219	0.277
Rayleigh distribution	0.175	0.215
Median-based	0.210	0.253
Global valley	0.099	0.281
MET	0.096	0.211
Kapur	0.129	0.215
Semivariance	0.050	0.105
Proposed	0.019	0.032

As shown in Table 1, the average *ME* of the proposed method (0.019) is the lowest among the 11 methods, followed in order by semivariance (0.050), valley-emphasis (0.078), MET(0.096), global valley (0.099), Kapur (0.129), and neighborhood valley-emphasis (0.139). The rest of the methods have average *ME* larger than 0.15. Based-on average *ME*, the proposed method is about 9 times better than the original Otsu method, and about 4 times better than the valley-emphasis method. The results show that the proposed weighting scheme which incorporates valley deepness measure with probability of occurrence is superior to the valley-emphasis weight. In addition, the proposed method has the lowest standard deviation, indicating that the method is robust across different image types. The second best method, the semivariance method, performs very well on the dataset. However, the computation cost of semivariance method is many folds higher than that of the Otsu method and the proposed method. The global valley method also has good thresholding results on most test images. But this method fails when the gray-level distributions do not have clear valleys. The results suggest that valley-point information alone is not sufficient for optimal threshold selection. It is interesting to see that some extended versions of the Otsu method do not perform as well on the dataset, such as the minimum variance method and the median-based method. These methods might work on some particular images but on average their performances are no better than the original Otsu method. In summary, the proposed weighting scheme greatly improves the accuracy and stability of the original Otsu method.

4 Conclusions

In this study, we revised the valley-emphasis weighting scheme and proposed a new weighting scheme to improve Otsu method for automatic threshold selection. The proposed weighting scheme incorporates a measure of valley deepness with probability of occurrence in the weight calculation to enhance the weighting effect. The performance of the proposed method was tested and compared with ten other popular thresholding methods using real-world images with ground truth. Experimental results show that the proposed method gives the best performance on the dataset. Based on average misclassification error, the proposed method is about 9 times better than the original Otsu method, and about 4 times better than the valley-emphasis method. The results also indicate that the proposed method is highly competitive with other widely used thresholding methods. Future research will extend the proposed method to multiple thresholding applications, and explore the possibility of applying the proposed weighting scheme on other threshold selection methods.

5 Acknowledgement

The authors would like to thank Dr. M. Sezgin [17] for the permission to use the test images and their ground truth. This research was partially supported by the Ministry of Science and Technology of the Republic of China under Grant MOST 104-2221-E-035-008-.

References

[1] M. Beauchemin, "Image thresholding based on semivariance," *Pattern Recognition Letters*, Vol. 34, pp. 456–462, 2013.

[2] E. Davies, "Stable bi-level and multi-level thresholding of images using a new global transformation," *IET Computer Vision*, Vol. 2, pp.60–74, 2008.

[3] C. Glasbey, "An analysis of histogram-based thresholding algorithms," *CVGIP: Graphical Models and Image Processing*, Vol. 55, pp. 532–537, 1993.

[4] J. Fan, B. Lei, "A modified valley-emphasis method for automatic thresholding," *Pattern Recognition Letters*, Vol. 33, pp. 703–708, 2012.

[5] Y. Liu, J. Xue, H. Li, "The study on the image thresholding segmentation algorithm," in *Proceedings of International Conference on Intelligent Systems Research and Mechatronics Engineering (ISRME 2015*), pp. 2306–2310, 2015.

[6] Z. Hou, Q. Hu, W. Nowinski, "On minimum variance thresholding," *Pattern Recognition Letters*, Vol. 27, pp. 1732–1743, 2006.

[7] J. Kapur, P. Sahoo, A. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphic and Image Processing*, Vol. 29, pp. 273–285, 1985.

[8] J. Kittler, J. Illingworth, "Minimum error thresholding," *Pattern Recognition*, Vol. 19, pp. 41–47, 1986.

[9] S. Lee, S. Chung, R. Park, "A comparative performance study of several global thresholding techniques for segmentation," *Computer Vision, Graphics and Image Processing*, Vol. 52, pp. 171–190, 1990.

[10] P. Liao, T. Chen, P. Chung, "A fast algorithm for multilevel thresholding," *Journal of Information Science and Engineering*, Vol. 17, pp. 713–727, 2001.

[11] T. Newman, "A survey of automated visual inspection," *Computer Vision and Image Understanding*, Vol. 61, pp. 231–262, 1995.

[12] H. Ng, "Automatic thresholding for defect detection," *Pattern Recognition Letters*, Vol. 27, pp. 1644–1649, 2006.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems Man Cybernet*, Vol. SMC-9, pp. 62–66, 1979.

[14] P. Pai, C. Chang, Y. Khan, M. Tsai, "An adaptable threshold detector," *Information Sciences*, Vol. 181, pp. 1463-1483, 2011.

[15] Y. Qiao, Q. Hu, G. Qian , S. Luo, W. Nowinski, "Thresholding based on variance and intensity contrast," *Pattern Recognition*, Vol. 40, pp. 596–608, 2007.

[16] P. Sahoo, S. Soltani, A. Wong, Y. Chan, "A survey of thresholding techniques," *Computer Vision, Graphics and Image Processing*, Vol. 41, pp. 233–260, 1988.

[17] M. Sezgin, B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, Vol. 13, pp. 146–156, 2004.

[18] N. Sharma, L. Aggarwal1, "Automated medical image segmentation techniques," Journal of Medical Physics, Vol. 35, pp. 3-14, 2010.

[19] Y. Wan, J. Wang, X. Sun, M. Hao, "A modified Otsu image segment method based on the Rayleigh distribution," in *Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology*, pp. 281-285, 2010.

[20] X. Xu, S. Xu, L. Jin, E. Song, "Characteristic analysis of Otsu threshold and its applications," *Pattern Recognition Letters*, Vol. 32, pp. 956-961, 2011.

[21] J. Xue, D. Titterington, "Median-based image thresholding," *Image and Vision Computing*, Vol. 29, pp. 361-367, 2011.

[22] W. Yasnoff, J. Mui, "Error measure for scene segmentation," *Pattern Recognition*, Vol. 9, pp. 217–231, 1977.

[23] Y. Zou, L. Fang, F. Dong, B. Lei, S. Sun, T. Jiang, P. Chen, "Median-based thresholding, minimum error thresholding, and their relationships with histogram-based image similarity," in *Proceedings of Sixth International Conference on Digital Image Processing (ICDIP 2014)*, doi:10.1117/12.2064335, 2014.