Performance Evaluation of Binarizations of Scanned Insect Footprints

Young W. Woo

Division of Computer & Visual Engineering
College of Engineering, Dongeui University
San 24, Gaya-Dong, Pusanjin-Gu, Pusan, 614-714, Korea

Abstract. The paper compares six conventional binarization methods for the special purpose of subsequent analysis of scanned insect footprints. We introduce a new performance criterion for performance evaluation. The six different binarization methods are selected from different methodologically categories, and the proposed performance criterion is related to the specific characteristics of insect footprints of having a very small percentage of object areas. The results indicate that a higher-order entropy binarization algorithm, such as proposed by Abutaleb, offers best results for further pattern recognition application steps for the analysis of scanned insect footprints.

Keyword: binarization, insect footprints, pattern recognition.

1 Introduction

In order to obtain accurate pattern recognition results using binarized images, it is obviously important to choose a suitable binarization algorithm. There exists a wide variety of binarization algorithms [5], and there is no “generally best” binarization algorithm for all kinds of gray level images. Some binarization algorithms are good for certain types of grey images but bad for other types of grey images. The project of interest (at CITR Auckland) is the classification of insects based on scanned tracks. Insects walk across a preinked card (as produced by Connovation Ltd., New Zealand), and leave a track on a white card which will be scanned on a standard flatbed scanner. The generated pictures are of very large scale due to the required resolution. Applications are related to, for example, pest control, environment protection, or monitoring of insect numbers.

We compared six representative binarization algorithms for binarizing scanned insect footprints, which have a relatively small object area compared to the non-object area, and selected the best binarization algorithm for pattern recognition for this particular application using a new performance criterion. The proposed binarization performance criterion is based on characteristics of insect footprints. In Section 2, we detail methods and formulas of six binarization algorithms. In Section 3, the proposed binarization performance criterion is presented. In Section 4, test images for experiments are shown and experimental results are presented. Finally we provide conclusions in Section 5.

2 Six Binarization Algorithms

Research on binarization of gray images dates back for more than thirty years. The paper [5] compares 40 binarization algorithms by two kinds of test data set. The first test data set is the set of 40 NDT (Non Destructive Testing) images, and the second test data set is the set of 40 document images. Binarization evaluation ranking tables in [5] show that the performance of each algorithm is different when different types of test data sets are applied. Only the binarization algorithm by Kittler and Illingworth [4] is best for both kinds of test data sets. So, as a first conclusion we know that conventional binarization algorithms' performance is highly depend on the or kind of gray images used in a particular application. In our case we have to binarize scanned insect footprints. We selected six different binarization algorithms that behaved "relatively good" for the two kinds of test data set in [5]. The selected six algorithms are as follows:

- Rosenfeld's convex hull binarization algorithm [7],
- Otsu's clustering binarization algorithm [6],
- Kittler and Illingworth's minimum error binarization algorithm [4],
- Kapur's entropic binarization algorithm [3],
- Abutaleb's higher-order entropy binarization algorithm [1], and
- Bernsen's local contrast binarization algorithm [2].

For a brief explanation of each algorithm, we use the following notation. The histogram and the probability mass function (PMF) of the image are indicated, respectively, by $h(g)$ and $p(g)$, for $g = 0 \ldots G_{max}$, where G_{max} is the maximum gray level in the image (which is typically 255). If the gray level range is not explicitly limited to a subinterval $[g_{\min}, g_{\max}]$, it will be assumed to be from 0 to G_{max}. The cumulative probability function is defined as

$$P(g) = \sum_{i=0}^{g} p(i) \tag{1}$$

The object (or *foreground*) and non-object (or *background*) PMFs are expressed as $P_f(g)$, for $0 \leq g \leq T$, and $P_b(g)$, for $T + 1 \leq g \leq G$, respectively, where T is the threshold value. The object and non-object area probabilities are calculated as follows:

$$P_f(T) = P_f = \sum_{g=0}^{T} p(g) \quad \text{and} \quad P_b(T) = P_b = \sum_{g=T+1}^{G} p(g) \tag{2}$$

The Shannon entropy, parametrically dependent on the threshold value T for foreground and background, is formulated as follows:

$$H_f(T) = -\sum_{g=0}^{T} p_f(g) \log p_f(g) \quad \text{and} \quad H_b(T) = -\sum_{g=T+1}^{G} p_b(g) \log p_b(g) \tag{3}$$

The mean and variance of the foreground and background as functions of the thresholding level T are denoted as follows:

$$m_f(T) = \sum_{g=0}^{T} g \cdot p(g) \quad \text{and} \quad \sigma_f^2(T) = \sum_{g=0}^{T} [g - m_f(T)]^2 p(g)$$
$$m_b(T) = \sum_{g=T+1}^{G} g \cdot p(g) \quad \text{and} \quad \sigma_b^2(T) = \sum_{g=T+1}^{G} [g - m_b(T)]^2 p(g) \qquad (4)$$

2.1 Rosenfeld's convex hull binarization algorithm

This algorithm is based on analyzing the concavities of the histogram $h(g)$ defined by its convex hull, $H(g)$; that is the set-theoretic difference $|H(g) - p(g)|$. When the convex hull of the histogram is calculated, the "deepest" concavity points become candidates for a threshold. In case of competing concavities, some object attribute feedback, such as low busyness of the edges of the thresholded image, can be used to select one of them. In this algorithm, the following equation is used for finding an optimal threshold value:

$$T_{opt} = \arg\max\{[p(g) - H(g)]\} \qquad (5)$$

2.2 Otsu's clustering binarization algorithm

This algorithm is to minimize the weighted sum of within-class variances of the foreground and background pixels to establish an optimum threshold. Recall that minimization of within-class variances is tantamount to the maximization of between-class scatter. This method gives satisfactory results when the numbers of pixels in each class are close to each other. The Otsu's algorithm is today one of the most referenced binarization algorithms. In this algorithm, the following equation is used for finding optimal threshold value:

$$T_{opt} = \arg\max\{\frac{P(T)[1 - P(T)][m_f(T) - m_b(T)]^2}{P(T)\sigma_f^2(T) + [1 - P(T)]\sigma_b^2(T)}\} \qquad (6)$$

2.3 Kittler and Illingworth's minimum error binarization algorithm

This algorithm assumes that the image can be characterized by a mixture distribution of foreground and background pixels: $p(g) = P(T) \cdot p_f(g) \quad + [1 - P(T)] \cdot p_b(g)$. Kittler and Illingworth's algorithm does not need the assumption that the foreground and background distribution function is an equal variance Gaussian density function and, in essence, addresses a minimum error Gaussian density-fitting problem. In this algorithm, the following equation is used for finding optimal threshold value:

$$\begin{aligned} T_{opt} = \arg\min\{&P(T)\log\sigma_f(T) + [1 - P(T)]\log\sigma_b(T) \\ &-P(T)\log P(T) - [1 - P(T)]\log[1 - P(T)]\} \end{aligned} \qquad (7)$$

where $\{\sigma_f(T),\ \sigma_b(T)\}$ are foreground and background standard deviations.

2.4 Kapur's entropic binarization algorithm

This algorithm assumes the image foreground and background as two different signal sources, so that when the sum of the two class entropies reaches its maximum, the image is said to be optimally binarized. In this algorithm, the following equation is used for finding an optimal threshold value:

$$T_{opt} = \arg\max[H_f(T) + H_b(T)] \text{ with} \tag{8}$$

$$H_f(T) = -\sum_{g=0}^{T} \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)} \quad \text{and} \quad H_b(T) = -\sum_{g=T+1}^{G} \frac{p(g)}{P(T)} \log \frac{p(g)}{P(T)}$$

2.5 Abutaleb's higher-order entropy binarization algorithm

This algorithm assumes the joint entropy of two related random variables, namely, the image gray level g at a pixel, and the average gray level $\bar{g}$ of a neighborhood centered at that pixel. Using the 2-D histogram $p(g,\ \bar{g})$, for any threshold pair $(T,\bar{T})$, one can calculate the cumulative distribution $P(T,\ \bar{T})$, and then define the foreground entropy as

$$H_f = -\sum_{i=1}^{T}\sum_{j=1}^{\bar{T}} \frac{p(g,\bar{g})}{P(T,\bar{T})} \log \frac{p(g,\bar{g})}{P(T,\bar{T})} \tag{9}$$

Similarly, one can define the background region's second order entropy. Under the assumption that the off-diagonal terms (i.e., the quadrants $[(0,\ T),\ (\bar{T}\ ,G)]$ and $[(T,\ G),\ (0,\ \bar{T})]$) are negligible and contain elements only due to image edges and noise, the optimal pair $(T,\ \bar{T})$ can be found as the minimizing value of the 2-D entropy functional. In this algorithm, the following equation is used for finding an optimal threshold value:

$$(T_{opt}, \bar{T}_{opt}) = \arg\min\{\log[P(T,\bar{T})[1-P(T,\bar{T})]] + H_f/P(T,\bar{T}) + H_b/[1-P(T,\bar{T})]\}$$

where $\quad H_f = -\sum\limits_{i=1}^{T}\sum\limits_{j=1}^{\bar{T}} \frac{p(g,\bar{g})}{P(T,\bar{T})} \log \frac{p(g,\bar{g})}{P(T,\bar{T})}$ and

$$H_b = -\sum_{i=T+1}^{G}\sum_{j=\bar{T}+1}^{G} \frac{p(g,\bar{g})}{1-P(T,\bar{T})} \log \frac{p(g,\bar{g})}{1-P(T,\bar{T})} \tag{10}$$

2.6 Bernsen's local contrast binarization algorithm

In the local binarization algorithm of Bernsen, the threshold is set at the midrange value, which is the mean of the minimum $I_{low}(i,j)$ and maximum $I_{high}(i,j)$ gray values in a local window of suggested size w =50. However, if the contrast $C(i,j) = I_{high}(i,j) - I_{low}(i,j)$ is below a certain threshold (this contrast threshold was 50), then that neighborhood is said to consist only of one class, print or background, depending on the value of $T(i,\ j)$. In this algorithm, the following equation is used for finding an optimal threshold value:

$$T(i,j) = 0.5\{\max_{w}[I(i+m,j+n)] + \min_{w}[I(i+m.j+n)]\} \tag{11}$$

where $w = 50$, provided contrast $C(i,j) = I_{high}(i,j) - I_{low}(i,j) \geq 50$.

In order to find an appropriate binarization algorithm for scanned insect footprints, the above 6 binarization algorithms have been implemented and their performances are evaluated using the proposed binarization performance criterion discussed in the next section.

3 The Proposed Binarization Performance Criterion

There are various conventional performance criteria for evaluation of binarization algorithms. In [5], the following five performance criteria are used for evaluating conventional binarization algorithms: misclassification error (ME), edge mismatch (EMM), relative foreground area error (RAE), modified Hausdorff distance (MHD), and region nonuniformity (NU). Four of these criteria except the last one need ground-truth images for comparison. In case of scanned insect footprints, it is almost impossible to acquire ground-truth images because nobody can decide easily which area is foreground or background. Because of this difficulty, we decided to use the last criterion, region nonuniformity(NU). The criterion NU measure is defined as follows:

$$NU = \frac{|F_T|}{|F_T + B_T|} \frac{\sigma_f^2}{\sigma^2} \tag{12}$$

where σ^2 represents the variance of the whole image, σ_f^2 represents the foreground variance, F_T, and B_T denote the areas of foreground and background

Fig. 1. A sample image.

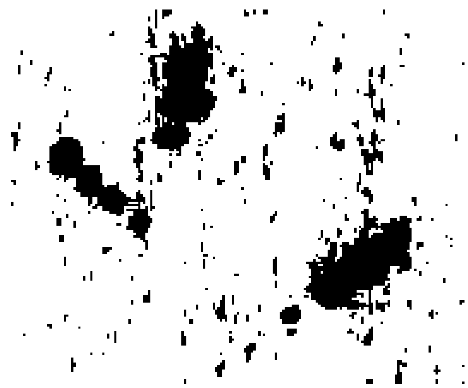

Fig. 2. The binarized image of Figure 1 for threshold value = 230.

pixels in the test image, and |.| is the cardinality of the set. It is expected that a well-segmented image will have a nonuniformity measure close to 0, while the worst case of NU=1 corresponds to an image for which background and foreground are indistinguishable up to second order moments [5]. However we found that there is a problem when this criterion is applied to scanned insect footprints' binarizations. The problem is that the value of this criterion goes nearer and nearer to 0 when a threshold value goes lower and lower. To show this result we have choose some area from a randomly selected insect footprint. The chosen area is shown as Figure 1. The binarized images at several threshold values are shown in Figure 2 to Figure 4.

We evaluated the NU values using the sample image of Figure 1, varying threshold values from 130 to 230. The result is shown in Table 1. In this table, we can see that the threshold value goes lower, the NU value becomes smaller, or,

Fig. 3. The binarized image of Figure 1 for threshold value = 180.

Fig. 4. The binarized image of Figure 1 for threshold value = 130.

in other words, "better". But when we closely look at the above three binarized images, Figure 3 is better than Figure 4. So the performance criterion of NU could not give an appropriate result on the scanned insect footprints.

We propose a new binarization performance criterion named *Minimum Number of Foreground Segments* (MNFS). In case of scanned insect footprints, it is important that foreground footprint pixels are given value 0 (black) everywhere where possible, and at the same time background noise area pixels are set to 255 (white) everywhere where possible. In Figure 2, it is decided that too many pixels to be 0, so many background noise area pixels are converted to 0 (i.e., foreground spots). In contrast, it is decided that too few pixels are set to 0, so many foreground spots' area pixels are converted to 255 (background area) in Figure 4. In these results, we can assume that "good" binarized images have a

Table 1. NU values for the sample image of Figure 1.

Threshold	NU
230	0.521458
220	0.227144
210	0.153669
200	0.103933
190	0.068894
180	0.044197
170	0.025973
160	0.013889
150	0.006888
140	0.003250
130	0.001024

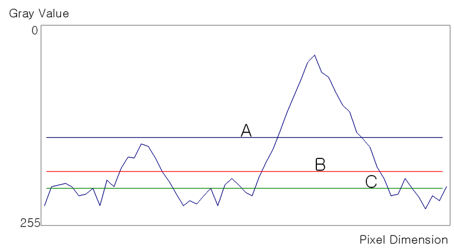

Fig. 5. Three cases of threshold value for binarization.

small number of disconnected regions but a large number of foreground pixels to be converted to 0 (foreground). We illustrate the concept of this idea in Figure 5.

In Figure 5, B is the assumed optimal threshold line that can produce the best binarized image because the only two "real spots" (one is a "clear spot" but another is a "dim spot") are converted to foreground objects. A is the threshold line that misses the dim spot because the threshold value is too low. C is the threshold line that converts lots of background noise area to foreground objects because the threshold value is too high. In case of the sample image of Figure 1, A means Figure 4, B means Figure 3, and C means Figure 2.

How can we choose a threshold value like B? The key idea for the choice of an optimal threshold value is to compare the number of disconnected segments (NDS) with the number of foreground pixels ($|F_T|$) and to compare the variance of background area. If the ratio of the two numbers ($NDS/|F_T|$) is smaller (i.e., threshold value becomes lower than A) and the normalized variance of background area ($\sigma_b^2(T)/\sigma^2(T)$) is smaller (i.e., threshold value becomes larger than

Table 2. The MNFS values for the sample image of Figure 1

.

Threshold	MNFS
230	0.00289038
220	0.000359181
218*	0.000342245
210	0.000554109
200	0.000790112
190	0.00146493
180	0.00299014
170	0.00541835
160	0.00717066
150	0.0113167
140	0.018377
130	0.0265409

Fig. 6. The binarized image of Figure 1 by threshold value = 218.

C), the threshold value can be considered to be "better". So we can consider the binarization as the best for binarization of scanned insect footprints if we achieve a minimum value of the product of ratio and normalized variance. The proposed binarization performance criterion is defined as follows:

$$MNFS = \frac{NDS}{|F_T|} \cdot \frac{\sigma_b^2(T)}{\sigma^2(T)} \quad (13)$$

where NDS indicates the number of disconnected foreground segments. For example, let us evaluate the MNFS values using the sample image of Figure 1, varying threshold values from 130 to 230. The result is shown in Table 2. In this sample image, the threshold value of 218 has the minimum MNFS value. The binarized image of the sample image of Figure 1 using threshold value of 218, is shown in Figure 6.

4 Test Images and Experimental Results

Our test images consisted of a variety of 16 images of American Cockroach, 30 images of Black Cockroach, and 25 images of Native Bush Cockroach. All images are scanned by 1200 DPI in 8-bit gray image format. Several test images are shown in Figure 7. The two images on the left are American Cockroaches, the two images in the middle are Black Cockroaches, and the two images on the right are Native Bush Cockroaches.

Results are shown in Figures 8, 9 and 10 for the sample image of Figure 1. A number in parenthesis in the caption is the calculated global threshold value for the algorithm. Rosenfeld's, Abutaleb's, and Bernsen's algorithms do not have global threshold values because they are locally adaptive algorithms.

The average MNFS values are given in Table 3 using 71 test images. In several test images, Otsu's and Kittler's algorithms result in incorrect threshold values

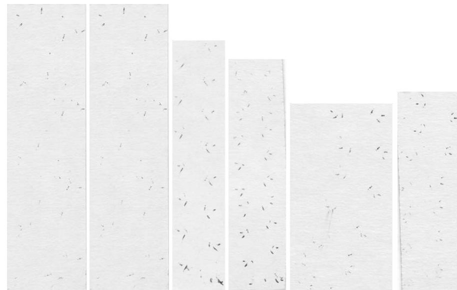

Fig. 7. Six sample images for testing.

Fig. 8. Binarized images using Rosenfeld's (left) and Otsu's [198] algorithm.

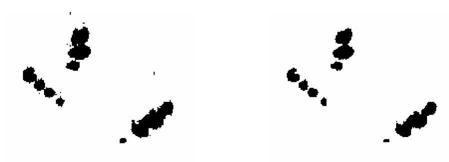

Fig. 9. Binarized images using Kittler's (left) [222] and Kapur's [214] algorithm.

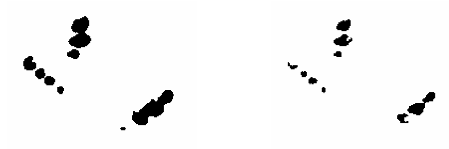

Fig. 10. Binarized images using Abutaleb's (left) and Bernsen's algorithm.

Table 3. Average MNFS values.

Algorithm	Rosenfeld	Otsu	Kittler	Kapur	Abutaleb	Bernsen
MNFS	0.005607	0.004225	0.005145	0.004982	0.004034	0.007406

Fig. 11. Binarized test image using Abutaleb's algorithm.

(often too large threshold values). Otsu's algorithm produced incorrect threshold values on 6 test images, Kittler's algorithm produced incorrect threshold values on 13 test images. The MNFS values shown in Table 3 correspond to these incorrectly binarized images. We concluded that Otsu's and Kittler's algorithms are inadequate regardless of the MNFS value, and Abutaleb's algorithm is (in general) best for binarization of scanned insect footprints. A binarized test image using Abutaleb's algorithm is shown in Figure 11. An incorrectly binarized test image using Otsu's algorithm is shown in Figure 12, and an incorrectly binarized test image using Kittler's algorithm is shown in Figure 13.

5 Conclusions

We compared six different binarization algorithms and proposed a new binarization performance criterion to analyze the best performance for scanned insect footprints. The experimental results showed that Abutaleb's binarization method

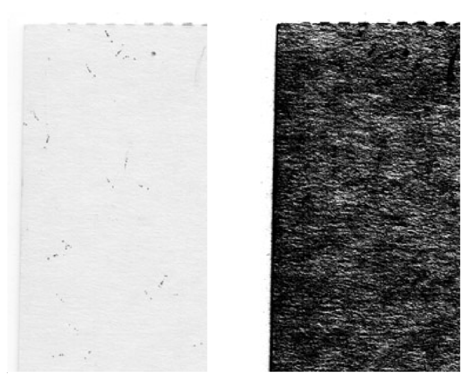

Fig. 12. Incorrectly binarized test image using Otsu's algorithm.

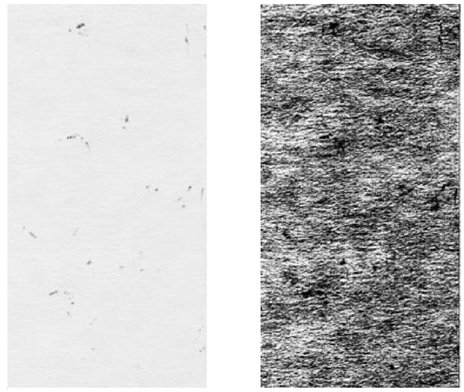

Fig. 13. Incorrectly binarized test image using Kittler's algorithm.

based on higher-order entropy produced (in general) the best binarized images. Binarized footprints have been further used in projects at CITR for property calculation, geometric modelling, and towards insect recognition. Results will be reported in forthcoming reports.

Acknowledgment: The reported binarization evaluation is part of an insect track recognition project at CITR, initiated and supported by Connovation Ltd., Auckland. The support of CITR, especially by Reinhard Klette, and of Connovation Ltd., especially by Warren Agnew, is very much appreciated by the author during his sabbatical stay at CITR.

References

1. A.S. Abutaleb. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision Graphics Image Processing*, **47**: 22–32, 1989.
2. J. Bernsen. Dynamic thresholding of gray level images. In Proc. of ICPR'86, pages 1251–1255, 1986.
3. 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. *Graphics Models Image Processing*, **29**: 273–285, 1985.
4. J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, **19**: 41–47, 1986.
5. M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* **13**: 146-165, Jan, 2004.
6. N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. on System, Man and Cybernetics* **SMC-9**: 62–66, 1979.
7. A. Rosenfeld and P. De la Torre. Histogram concavity analysis as an aid in threshold selection. *IEEE Trans. on System, Man and Cybernetics* **SMC-13**: 231–235, 1983.