Comparing Threshold-Selection Methods for Image Segmentation: Application to Defect Detection in Automated Visual Inspection Systems

Rafael López-Leyva, Alfonso Rojas-Domínguez[(✉)],
Juan Pablo Flores-Mendoza, Miguel Ángel Casillas-Araiza,
and Raúl Santiago-Montero

Tecnológico Nacional de México - Instituto Tecnológico de León,
Av. Tecnológico S/N - Frac. Industrial Julián de Obregón,
37290 Leóng Gto., Mexico
alfonso.rojas@gmail.com

Abstract. Automatic detection of defects on the surface of products or raw material is an important task in the field of automated visual inspection. Thresholding is a method for image segmentation that is often used for the detection of said defects. Several methods to select the optimal thresholding values automatically on a per-image base have been described in the literature. Some of these are particularly designed to deal with mostly homogeneous images such as those of product surfaces with some defects, but have not been tested sufficiently and not in the context of automated visual inspection. In this work we present a comparison based on such experimental conditions by means of an automated visual inspection station and a set of images specially acquired for this purpose. The methods that were compared are: the Otsu's method, the Valley-Emphasis method, the Valley Emphasis with Neighborhood method, the Kittler-Illingworth's method, and the Maximum Similarity Thresholding method. The highest performance, with a statistically significant difference, was obtained by Maximum Similarity Thresholding.

Keywords: Automated thresholding · Otsu · Valley-Emphasis · Maximum Similarity Thresholding · Kittler-Illingworth

1 Introduction

Thresholding is a segmentation technique widely used in industrial applications. It is used when there is a clear difference between the objects and the scene background. Therefore, the scene should be characterized by a uniform background, similarity between the objects to be segmented and homogeneous appearance of these; otherwise all the pixels that compose the scene could be assigned to a single class. In industrial applications

usually the illumination and other scene properties can be easily controlled. Thus, the main problem for automated segmentation via thresholding lies in determining the threshold value that segments the image with optimal results.

Several methods designed to find the optimal thresholding value have been described in the literature; the most popular is Otsu's method [1]. This method selects the threshold value that maximizes the variance between classes of an image histogram and for this reason it is most successful when the images show a bimodal or close to bimodal histogram (in the case of two output classes) or multi modal (in the cases of multiple output classes), but the method usually fails when the histogram is unimodal or near to be unimodal. The latter situation occurs frequently in visual inspection of surfaces, where the test images are mostly uniform and the defects are small and may be represented by grey levels that are similar to those of the image background. Several methods have been proposed to reduce the limitations of Otsu's method, such as the Valley-Emphasis method [2], the Valley-Emphasis using Neighbourhood method [3], the Kittler-Illingworth's method [4] and the MST method [5]. However, although these studies have proposed valuable improvements, there is a lack of experimental results from real-life scenarios to validate the proposals. The need to objectively assess the capabilities of different thresholding methods in order to use them in Automated Visual Inspection (AVI) systems, motivated this study.

For this work we have implement an Automated Visual Inspection System (AVI) [9] consisting of an inspection point equipped with industrial-grade lighting and image acquisition equipment on a conveyor belt. Using this equipment, a set of images designed to test the different thresholding methods were produced, including annotations made by manually selecting a threshold value. For the comparative evaluation two similarity measures were employed: the Tanimoto Coefficient [6, 7] and the Pearson Correlation Coefficient (PCC) [5, 8]. Although our test conditions are not those of an industrial environment, they are sufficiently similar to real-life scenarios as to provide the required evidence in favour of a thresholding method.

The rest of this article is organized as follows: Sect. 2 summarized the thresholding methods under comparison; Sect. 3 describes the implementation of our AVI system and experimental design. The experimental results are contained in Sect. 4. The analysis and discussion of the results are found in Sect. 5 and finally the conclusions and directions of future work are provided in Sect. 6.

2   Theoretical Background

In this section we review a number of methods for automatic selection of threshold values; the methods that we will discuss are: the Otsu's method, the Valley-Emphasis method, the Valley-Emphasis method using Neighborhood, the Kittler-Illingworth's method and the MST method. For a more general discussion regarding thresholding techniques *cf.* [10].

2.1 The Otsu's Method

An image is a bidimensional matrix of gray intensity levels that contains N pixels with gray level values between 0 and 255 (1, L). The number of pixels with a gray level i is denoted as f_i, and the probability of occurrence of gray level i is given by:

$$P_i = \frac{f_i}{N} \tag{1}$$

The total average of gray level in the image can be calculated in this way:

$$\mu_T = \sum_{i=0}^{L-1} i p_i \tag{2}$$

By segmenting the image using a single threshold we get two disjoint regions C_1 and C_2, which are formed by the area of pixels with gray levels $[1,\ldots,t]$ and $[t+1,\ldots,L]$ respectively: Normally these classes correspond to the object of interest and the background of the image. Then the probability distributions of the gray level for these two classes are:

$$C_1 = \frac{P_i}{\omega_1(t)},\ldots,\frac{P_t}{\omega_1(t)} \quad \text{and} \quad C_2 = \frac{P_{t+1}}{\omega_2(t)},\frac{P_{t+2}}{\omega_2(t)},\ldots,\frac{P_L}{\omega_2(t)} \tag{3}$$

Where the probabilities for the two classes are calculated as follows:

$$\omega_1(t) = \sum_{i=1}^{t} P_i \quad \text{and} \quad \omega_2(t) = \sum_{i=t+1}^{L} P_i \tag{4}$$

The mean values of the gray levels for the two classes are calculated as follows:

$$\mu_1 = \sum_{i=1}^{t} \frac{i \cdot P_i}{\omega_1(t)} \quad \text{and} \quad \mu_2 = \sum_{i=t+1}^{L} \frac{i \cdot P_i}{\omega_2(t)} \tag{5}$$

In Otsu's method [1] the optimal threshold t^* is determined by maximizing the variance between classes, which is denoted in the following objective function:

$$t^* = \frac{ArgMax}{0 \le t < L} \{\sigma_B^2(t)\} \tag{6}$$

Where the variance between classes σ_B^2 is defined by:

$$\sigma_B^2 = \omega_1 \cdot (\mu_1 - \mu_T)^2 + \omega_2 \cdot (\mu_2 - \mu_T)^2 \tag{7}$$

Otsu's method is easy to calculate, but it only works properly for images containing bimodal distribution histograms, which means that the object and background have comparable gray level variances, but fails to find the optimal threshold when the histogram contains a unimodal distribution or close to unimodal. Following this observation, the Valley-Emphasis method [2] was proposed. This is discussed below.

2.2   Valley-Emphasis Method

The Valley-Emphasis method [2] is the result of observing that the Otsu's method does not behave as desired with (near to) unimodal distribution histograms. The emphasis of valley method selects threshold values that have a small probability of occurrence, while at the same time maximizes the variance between groups as the Otsu's method does. This modification means changing the target function of the Otsu's method (Eq. 5) by applying the weight function $(1 - P_t)$:

$$t^* = \frac{ArgMax}{0 \leq t < L} \left\{ (1 - P_t)\left(\omega_1(t)\mu_1^2(t) + \omega_2(t)\mu_2^2(t)\right) \right\} \tag{8}$$

The smaller the value P_t (probability of the threshold value t), the greater the result of (8) since it is multiplied by the complement of this probability. This weight function ensures that the result will always be a threshold value that is in a "valley" or bottom edge of the gray level distribution. The results reported in [2] show that the Valley-Emphasis method produced a misclassification two orders of magnitude below that of the Otsu's method, but these results were not based on sufficient experimentation.

2.3   Valley-Emphasis Using Neighborhood Method

It has been argued that the Valley-Emphasis method cannot improve the quality of segmentation in some cases because, being based on a single gray level value, is not robust enough. A variation of the Valley Emphasis method, called Valley-Emphasis using Neighborhood has been proposed that attempts to alleviate this issue by considering the information contained in the neighborhood around the valley point. The improvement consists in using a modified weight function that is applied on $\sigma_B^2(t)$. This weight function is defined as follows. Using the histogram $\{h(i)\}$ of an image, the neighborhood $\overline{h}(i)$ of gray level i is:

$$\overline{h}(i) = [h(i - m) + \ldots + h(i - 1) + h(i) + h(i + 1) + \ldots + h(i + m)] \tag{9}$$

where m is the size of the neighborhood. The Valley-Emphasis method obtains an optimal threshold by modifying the objective function of the Otsu's method (Eq. 5) in the following way:

$$\xi(t) = \left(1 - \overline{h}(i)\right)\left(\omega_0(i)\mu_1^2(i) + \omega_1(i)\mu_2^2(i)\right) \tag{10}$$

The optimal threshold t^* is the gray level value that maximizes (11):

$$t^* = \frac{ArgMax}{0 \leq t < L} \xi(t) \tag{11}$$

The contribution of the Valley-Emphasis method with Neighborhood is that it considers the neighbors of each gray value in its search for the optimal threshold (this provides robustness) and that it preserves the idea that those with the smallest occurrence

must have the largest influence (the proposal of the Valley-Emphasis method). The results obtained in [3] suggest that a neighborhood of size 11 or slightly larger produces the best segmentation results.

2.4 The Kittler–Illingworth's Method

The main idea behind the method called Kittler-Illingworth thresholding is to directly optimize (minimize) the average rate of misclassification. The method obtains the optimal threshold using the minimum thresholding error in calculations based on their objective function:

$$J(t) = 1 + 2[P_1(t)log\sigma_1(t) + P_2(t)log\sigma_2(t)] - 2[P_1(t)logP_1(t) + P_2(t)logP_2(t)] \tag{12}$$

where $\sigma_i^2(t) = \left[\sum_{i=1}^{L-1}\{i - \mu_i(t)\}^2 f_i\right]/P_i(t)$. After obtaining the target values for the gray levels of the image, the method search the minimum value of the function, so that which is defined:

$$t^* = Min\{J(t)\} \tag{13}$$

The results obtained in [4] show that the method achieves a better result than the Otsu's and Ridler's [11] methods when the images are homogeneous or, in cases of bimodal histograms, when one of the peaks is large and the other one is small, meaning that the optimal threshold tends to fall where there is less likelihood of occurrence between the two modes.

2.5 Maximum Similarity Thresholding (MST) Method

The MST method selects the optimal threshold as the value that maximizes the similarity between the edge content of the thresholded image and the edge content of the original image computed as follows:

$$T(t) = \prod_{i=1}^{k} \|\nabla G(x, y, \sigma_i) * t\| \tag{14}$$

where $G(x, y, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x^2 + y^2)}{2\sigma_i^2}}$ and the symbols ∇ and $*$ represent the derivate and the convolution, respectively. The results obtained in [5] show that, using the Pearson's Correlation Coefficient as a similarity metric, the MST method on average achieves similarities two orders of magnitude larger than the Otsu's method and the Valley-Emphasis method. The disadvantage of this method, compared to other methods, is its greater complexity, which could affect its performance in real time applications.

3   Development

In an industrial environment, control of image capture and the lighting is extremely important because these produce uniformity in the images, thus simplifying inspection tasks. In this work, the following parameters were considered to guarantee said uniformity in the acquisition: exposure time, signal amplification, automatic brightness control, external control to manage lighting (strobe mode), and synchronizing image capture with the movement of the conveyor belt. A CCD digital camera for machine vision applications, the Grey Point Flea3 model [12], equipped with an 8 mm focal-length lens was used in the acquisition. The choice of the lens is related to the working distance (between the object on the conveyor belt and the camera). A ring-type LED lamp with white light was used in the experiments. Synchronization between frame acquisition and light strobe is achieved through a trigger signal produced by the camera (Fig. 1).
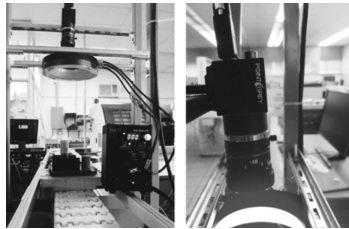

Fig. 1.   AVI installed on the conveyor band, which has a view camera and white lighting of ring type.

After synchronization of the camera with the strobe light and the movement of the conveyor belt, a number of images of a paper strip moving on the conveyor were obtained. The paper strip contained several defects, designed to simulate those defects that can be observed in the production of different types of fabrics and other homogeneous materials. The size of the defects range from 11 to 27133 pixels; the average intensity value of the defects is 161.8 ± 12.2 (using 8 bpp gray scale images) while the average intensity of the background is 225.7 ± 1.3; the mean intensity difference between defect and background is approximately 64. A sample of the images produced is shown in Fig. 2.
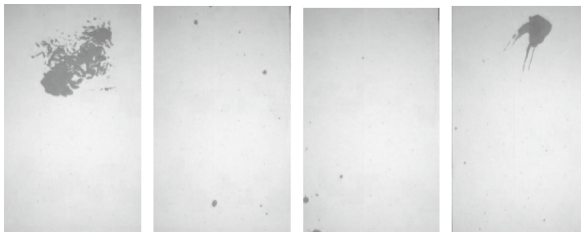

Fig. 2.   Different defects found in paper strips moving on the conveyor belt.

After selecting the frames that were deemed most useful to our purposes, 97 images were saved that most clearly show the defects on the paper strips. These constitute the set of test images on which the different threshold selection algorithms summarized in Sect. 2 are tested. An example of applying each of the methods under comparison on a test image is illustrated in Fig. 3.
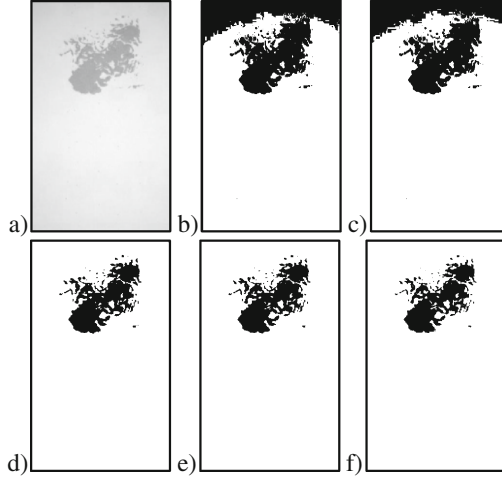

Fig. 3. Example of thresholding of an image with a defect to the paper, where (a) is the original image, (b) is the thresholding with Otsu ($t^* = 204$), (c) Valley-Emphasis ($t^* = 205$), (d) Valley-Emphasis using neighborhood ($t^* = 180$), (e) Kittler Illingworth ($t^* = 179$), (f) and the MST method ($t^* = 174$).

In order to quantify the quality of the results obtained by means of applying each of the threshold selection algorithms on our test images, the similarity between the original image and the thresholded images was computed. For this purpose, a set of binary images was produced by manually selecting the best threshold value in each case. The similarity between these reference images, referred to as Ground Truth masks, and each binary image produced by the thresholding algorithms, was measured using two measures; the Tanimoto's Coefficient (TC) [6, 7] and Pearson's Correlation Coefficient (PCC) [5, 8] which are defined as follows:

$$TC = \frac{N(A \cap B)}{N(A \cup B)} \tag{18}$$

$$PCC = \frac{\sum_{k=1}^{L}(i_k - \mu_A)(i_k - \mu_B)}{\sqrt{\sum_{k=1}^{L}\left(i_k - \mu_A^2\right)\sum_{k=1}^{L}\left(i_k - \mu_B^2\right)}} \tag{19}$$

where A and B denote the images under comparison, in this case the Ground Truth masks against the automatically segmented images.

4   Experimental Results

Using the Tanimoto's Coefficient [6, 7] as similarity measure, the results of the comparison performed with base on our test set are summarized in Table 1. The corresponding boxplots and the Wilcoxon's rank sum test of these results are shown in Fig. 4 and Table 3 respectively.

Table 1. Tanimoto coefficient between Ground Truth and automatic segmentation

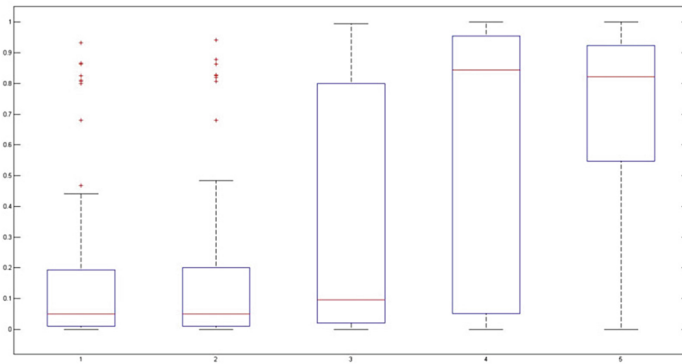|	Otsu	Valley-Emphasis	Valley-Emphasis neighborhood	Kittler-Illingworth	MST
Percent similarity range	0.003–0.932	0.003–0.940	0.005–0.995	0.000–1.000	0.000–1.000
Average	0.158	0.160	0.331	0.614	0.686
Standard deviation	0.232	0.234	0.381	0.406	0.304

Fig. 4. Boxplots of the Tanimoto's Coefficient between thresholded test images against the corresponding Ground Truth, from left to right: Otsu's method; Valley-Emphasis method; Valley-Emphasis using neighborhood; the Kittler-Illingworth's method and MST method.

Using instead the Pearson's Correlation Coefficient [8, 9] as similarity measure, the results summarized in Table 3 were obtained. The corresponding boxplots are shown in Fig. 5 and the Wilcoxon's rank sum test in Table 3 (Table 2).

Table 2. Pearson's coefficient between Ground Truth and automatic segmentation.

	Otsu	Valley-Emphasis	Valley-Emphasis neighborhood	Kittler-Illingworth	MST
Percent similarity range	0.015– 0.961	0.016–0.966	0.021–0.997	0.001–1.000	0.001–1.000
Average	0.281	0.284	0.442	0.679	0.775
Standard deviation	0.249	0.250	0.355	0.382	0.275

Figure 5 displays the results for each of the thresholding methods that were used:
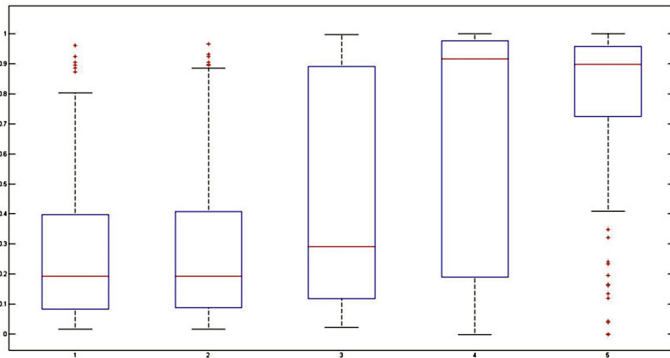

Fig. 5. Boxplot corresponding to the values resulting of similarity of Pearson's Correlation Coefficient between all thresholded images presented defects against the corresponding masks to each them, where (1) refers to Otsu's, (2) the Valley-Emphasis, (3) the Valley-Emphasis using Neighborhood, (4) the Kittler-Illingworth's and (5) the MST methods.

Table 3. Wilcoxon's rank sum test between TC of the methods and PCC of the methods.

Wilcoxon's rank sum test between Tanimoto's Coefficient of the methods									
Otsu VE	Otsu VEN	Otsu KI	Otsu MST	VE VEN	VE KI	VE MST	VEN KI	VEN MST	KI MST
0.895	0.005	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.967
Wilcoxon's rank sum test between PCC of the methods									
Otsu VE	Otsu VEN	Otsu KI	Otsu MST	VE VEN	VE KI	VE MST	VEN KI	VEN MST	KI MST
0.891	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.904

5 Discussion

The results obtained with the Tanimoto's Coefficient as the similarity measure show that Otsu's method, the Valley-Emphasis method and the Valley-Emphasis method using Neighborhood obtained the lowest values (based on the corresponding median values) with an average similarity against the Ground Truth between 0.15 to 0.33 and a standard deviation between 0.20 and 0.40 all of these methods appear to be inferior to the rest. The Kittler-Illingworth's method and the MST method obtained the highest performance values; however, with average values of 0.61 and 0.68 respectively and a standard deviation of approximately 0.4 and 0.3 respectively, the MST method appears to be superior. The statistical significance of these observations was validated by means of the Wilcoxon's rank sum test, with which we compared all methods against each other to

verify that the differences between them is statistically supported. In this test, the null hypothesis is that there is no statistical difference between the data samples being compared. When comparing the MST method against the other methods, the test indicates that the null hypothesis can be rejected (at a significance level of 0.05) in the case of Otsu's, Valley-Emphasis, and Valley-Emphasis using Neighborhood, since all of these produced a $P_{value} = 0.000$. Meanwhile, in the comparison against the Kittler-Illingworth's method, with a $P_{value} = 0.967$ the test shows that there is no evidence to reject the null hypothesis; in other words, there is no statistical difference between the results of the MST and the Kittler-Illingworth's method.

Regarding the results based on the Pearson's Correlation Coefficient as the similarity measure, we can observe that although the numerical values vary slightly, a very similar behavior as that when using the Tanimoto Coefficient is obtained. This observation was also verified by means of the Wilcoxon's rank sum test, with which we compared the methods against each other under the same hypotheses and significance level. With a $P_{value} = 0.904$ the null hypothesis cannot be rejected in the comparison between the MST method and the Kittler-Illingworth's, while in the comparison against the other methods there is evidence to reject it ($P_{value} = 0.000$).

6   Conclusion

In this work, several methods for automatic determination of the optimal threshold value for image segmentation were compared using a set of test images specially designed to test the utility of the algorithms to be employed in AVI systems. The test images depict varying defects in paper strips and were acquired under industrial-standard conditions. Tested under two different similarity measures (the Tanimoto's Coefficient and the Pearson's Correlation Coefficient), the MST method obtained the highest similarity against the other methods, with an average measure of TC = 0.68 and PCC = 0.77. As future work, we plan to employ the MST method in our AVI tasks and particularly in the automated detection of defects on homogeneous surfaces. The experiments presented provide us with the necessary knowledge to make this informed decision, which was not possible prior to the realization of this work.

References

1. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
2. Hui-Fuang, Ng: Automatic thresholding for defect detection. Pattern Recogn. Lett. **27**, 1644–1649 (2006)
3. Fan, J., Lei, B.: A modified valley-emphasis method for automatic thresholding. Pattern Recogn. Lett. **33**, 703–708 (2012)
4. Kittler, J., Illingworth, J.: Minimum error thresholding. Pattern Recogn. **19**, 41–47 (1986)
5. Yaobin, Z., Shuifa, S., Fangmin, D.: Maximum similarity thresholding. Digital Sig. Process. **28**, 120–135 (2014)

6. Crum, W., Camara, O., Hill, D.: Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans. Med. Imaging **25**(11), 1451–1461 (2006)
7. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
8. Rodgers, J., Nicewander, W.: Thirteen ways to look at the correlation coefficient. Am. Stat. **42**, 59–66 (1988)
9. Christian, D., Bernd, A., Carsten, G.: Industrial Image Processing. Springer, Heidelberg (2013)
10. Davies, D.: Machine Vision – Theory, Algorithms, Practicalities. Elsevier, Philadelphia (2005)
11. Ridler, T., Calvard, S.: Picture thresholding using an iterative selection method. IEEE Trans. Syst. Man Cybern. **8**, 630–632 (1978)
12. Point Grey: Register Reference for Point Grey Digital Cameras (2015). www.ptgrey.com