Soft Comput (2018) 22:4197–4203
https://doi.org/10.1007/s00500-017-2709-1

FOCUS

Automatic image thresholding using Otsu's method and entropy weighting scheme for surface defect detection

Mai Thanh Nhat Truong[1] · **Sanghoon Kim**[1]

Published online: 14 July 2017

Abstract Defect detection is one of the most important tasks and a challenging problem for industrial quality control. Among the available visual inspection techniques, automatic thresholding is a commonly used approach for defect detection because of the simplicity in terms of its implementation and computing. In this paper, we propose an automatic thresholding technique, which is an improvement in Otsu's method, using an entropy weighting scheme. The proposed method enables the detection of extremely small defect regions compared to the product surface area. Experimental results confirm the efficiency of the proposed system over other techniques.

Keywords Automatic thresholding · Nondestructive testing · Defect detection · Entropy

1 Introduction

Every day enormous volumes of goods are produced. This is a consequence of the availability of advanced technologies and the rapid growth of industrial production. Large-scale production has made manual defect detection infeasible because of the vast amount of production that must be inspected in a short time period. Therefore, research on machine-assisted defect detection systems has attracted a considerable amount of attention recently. The two primary types of inspection techniques are destructive and nondestructive testing. In destructive testing, the materials or products are dismantled to examine their physical properties such as strength, robustness, and durability. Destructive testing methods are also used for determining design weaknesses or revealing hidden defects that cannot be detected by other means. In contrast to destructive testing, nondestructive testing (NDT) is a group of inspection techniques used to examine the properties of articles or identify defects without altering the original characteristics or causing damage. Because products can function normally after being inspected, NDT is a preferable method that can save both time and money in defect detection. Common NDT methods are ultrasonic, thermographic, infrared thermography, visual or visual inspection, liquid penetrant, and magnetic particle testing (Gholizadeh 2016).

Among the common NDT methods, visual inspection is highly preferable because it is a high-speed process and does not require complex equipment. In this method, cameras, scanners, and sensors are used to create visual representations of the materials or products. Then, computer vision algorithms or imaging processing techniques are used to detect the defects from the acquired visual data. Several visual inspection methods have been published and applied successfully (Shankar and Zhong 2005; Ng 2006; Asha et al. 2012; Bhajantri et al. 2013; Uddin et al. 2014), demonstrating the potential of this research trend. Different algorithms can be applied for inspecting different kinds of product such as fabrics, paper, foils, or steel. However, the majority of these algorithms are high complexity (Tolba and Raafat 2015), implying that they require long execution time to complete the inspection process. Therefore, there is a requirement for fast and automatic algorithms that can produce acceptable results for several types of product without manual tuning.

Communicated by J. Park.

✉ Sanghoon Kim
kimsh@hknu.ac.kr

[1] Department of Electrical, Electronics, and Control Engineering, Hankyong National University, 327 Jungang-ro, Anseong-si, Gyeonggi-do 456-749, Republic of Korea

One of the promising approaches for developing such methods is automatic thresholding.

Automatic thresholding has been widely used not only in automated visual quality inspection systems, but also in many other areas of computer vision such as segmentation (Al-Tairi et al. 2014), edge detection (Verma et al. 2013), and movement analysis (Hussain et al. 2012). In this image analysis technique, an optimal threshold value is automatically selected for partitioning an image into a background and a set of objects of interest, that is, it separates objects from the background by converting grayscale images into k-ary images. Automatic thresholding is a simple and effective technique; for this reason, it has various applications in image processing. Because a considerable number of automatic thresholding methods have been developed, several surveys for well-known techniques have been conducted. Comprehensive reviews are presented in (Sezgin and Sankur 2004) and (Chaki et al 2014).

Otsu's method, a state-of-the-art automatic thresholding technique, is the basis for several automatic defect detection proposals. It determines optimal threshold values that maximize the between-class variances of the foreground and background. Studies demonstrate that Otsu's method is effective for thresholding a histogram with bimodal or multimodal distribution. Otsu's method fails if the histogram is unimodal or almost unimodal (Ng 2006). Therefore, it provides acceptable results for thresholding general real-life images and fails when being applied in visual inspection systems because of the nature of the input images. Real-life image histograms typically have multimodal distributions because the objects are usually large compared with the image size. This property enables Otsu's method to determine the optimal thresholds more easily. However, the histograms of images obtained from visual inspection systems generally have unimodal distributions because of the uniformity in luminance and size of the defect regions, which prevents Otsu's method from identifying the desired threshold values.

Several improvements have been proposed for the purpose of overcoming this drawback of Otsu's method in defect detection. Although these improvements have demonstrated their efficiency, they can fail in some cases, especially when detecting micro-sized defects. Moreover, some methods require parameters that must be assigned manually. For this reason, such methods are not robust when being applied in different inspection systems. In this research work, we propose a novel visual inspection algorithm that is an improvement to Otsu's method. Our proposal uses the entropy theory and does not require user input parameters. The algorithm is fully automatic and capable of detecting extremely small defect regions compared to the product surface area.

2 Otsu's method and its improvement

In this section, we provide a summary and brief analysis of Otsu's method (Otsu 1979), an automatic clustering-based image thresholding technique. Otsu's method remains one of the most frequently referenced thresholding techniques (Sezgin and Sankur 2004). We also introduce variants of Otsu's method that have been developed for defect detection.

Let $I = f(x, y)$ be a digital image, where f is an intensity function. The value of $f(x, y)$ is the gray value of the pixel at location (x, y), which lies in the interval $[0, L-1]$, where L is the number of gray levels. Let the number of pixels with gray value i be n_i and N be the total number of pixels in I. The histogram is considered a probability distribution:

$$p_i = \frac{n_i}{N} \tag{1}$$

Then, the average gray value of I is:

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

As a result of thresholding, the pixels of I are divided into two classes, $C_1 = \{(x, y) \mid 0 \le f(x, y) \le k\}$ and $C_2 = \{(x, y) \mid k+1 \le f(x, y) \le L-1\}$, where k is the selected threshold value. Normally, C_1 is the foreground and C_2 is the background, or vice versa. The probability of class occurrence and the average gray value of each class are, respectively, computed as

$$\omega_1(k) = \sum_{i=0}^{k} p_i, \quad \omega_2(k) = \sum_{i=k+1}^{L-1} p_i \tag{3}$$

and

$$\mu_1(k) = \sum_{i=0}^{k} \frac{i p_i}{\omega_1(k)}, \quad \mu_2(k) = \sum_{i=k+1}^{L-1} \frac{i p_i}{\omega_2(k)} \tag{4}$$

In Otsu's method, the performance of the result threshold is measured by considering the distinction between the foreground and background. Using this criterion, the optimal threshold k^* must maximize between-class variance:

$$k^* = \underset{0 \le k \le L-1}{\arg\max} \{\omega_1(k)(\mu_1(k) - \mu_I)^2 + \omega_2(k)(\mu_2(k) - \mu_I)^2\} \tag{5}$$

Another equivalent formula for calculating the optimal threshold is given in (Liao et al. 2001). This formula is simpler than the original:

$$k^* = \underset{0 \le k \le L-1}{\arg\max} \{\omega_1(k)\mu_1^2(k) + \omega_2(k)\mu_2^2(k)\} \tag{6}$$

Otsu's method produces acceptable results when objects in the images are large and sufficiently different from the background. More specifically, it works for images whose histograms have bimodal or multimodal distributions. The histograms of images from a visual inspection system, however, appear to be unimodal because the defect regions are small compared with the image size. This prevents Otsu's method from determining the optimal thresholds. Therefore, Otsu's method is not appropriate for defect detection.

Several improvements to Otsu's method for defect detection have been proposed. The basic idea is the addition of a weight W to the objective function (6) for adjusting the output threshold:

$$k^* = \underset{0\le k\le L-1}{\arg\max}\{W(\omega_1(k)\mu_1^2(k) + \omega_2(k)\mu_2^2(k))\} \tag{7}$$

The earliest weighting scheme for Otsu's method was the valley-emphasis method (Ng 2006):

$$k^* = \underset{0\le k\le L-1}{\arg\max}\{(1 - p_k)(\omega_1(k)\mu_1^2(k) + \omega_2(k)\mu_2^2(k))\} \tag{8}$$

where the weight W is $1 - p_k$. The key of the valley-emphasis method is the complement of probability of occurrence $(1 - p_k)$. The smaller the value of p_k (i.e., the lower the probability of occurrence of gray value k), the larger the weight becomes. This weight ensures that the output threshold will always be in the valley or bottom rim of the histogram, hence the name "valley-emphasis." Figure 1 indicates the efficiency of the valley-emphasis method over Otsu's method in defect detection. The selected thresholds are marked by vertical red lines. The original image (Fig. 1a) was acquired from (Ng 2006) through Rightslink®, and the results were produced by our implementation.

Following the proposal of the valley-emphasis method, a number of weighting schemes for Otsu's method that also use the probability of occurrence were proposed (Fan and Lei 2012; Ng et al. 2013). There is also a weighting scheme based on Fisher's linear discriminant analysis (LDA) (Liu et al. 2014). Figure 2 indicates thresholding results from our implementations of aforementioned methods. The original image (Fig. 2a) was acquired from (Liu et al. 2014), and it was published under Creative Commons Attribution License.

Even though these methods have significantly increased the accuracy of defect detection, as illustrated in Fig. 2, their performance is not stable. Moreover, in some methods, the user must choose appropriate parameters to obtain acceptable results. In the following section, we introduce a new weighting scheme to improve the performance of Otsu's method without the requirement of tuning or assigning parameters.

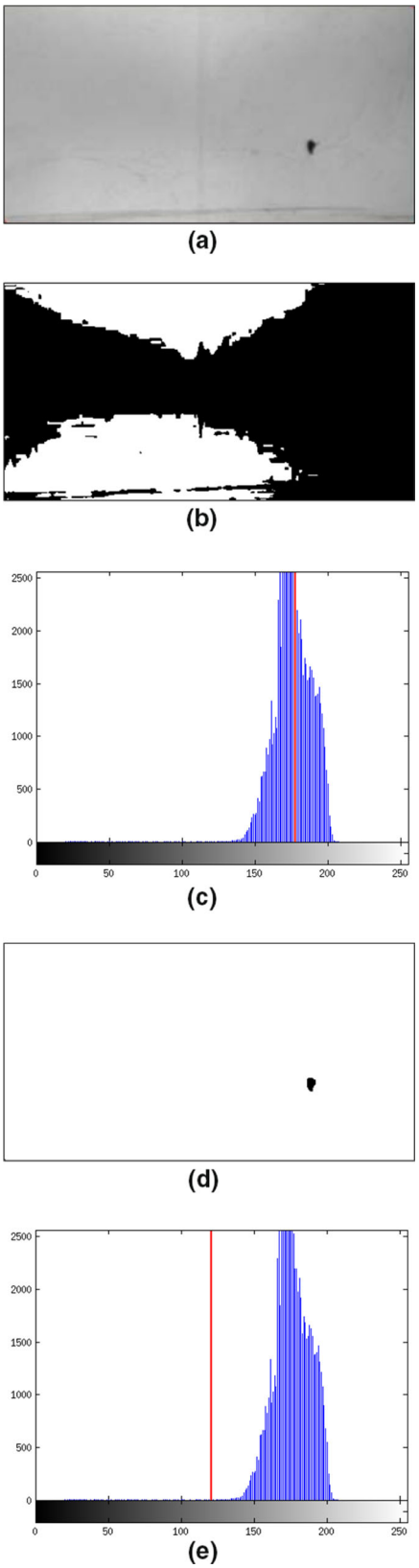

Fig. 1 Capability of the valley-emphasis method in defect detection. **a** Original image. **b** Result from Otsu's method. **c** Threshold from Otsu's method. **d** Result from valley-emphasis method. **e** Threshold from valley-emphasis method

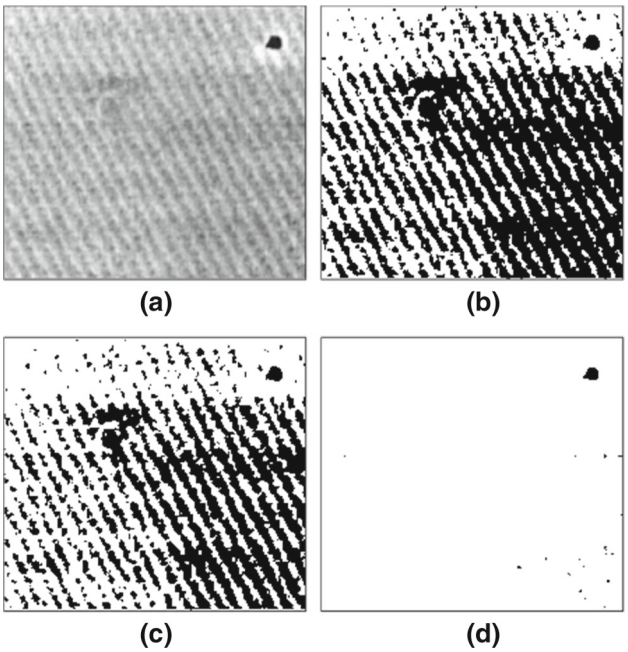

Fig. 2 Result comparison between thresholding methods in defect detection. **a** Original image. **b** Otsu's method. **c** Valley-emphasis. **d** Fisher's LDA

3 Improving Otsu's method using entropy

Image entropy is a statistical measure of randomness that can be used to represent the characteristics of an image. Low-entropy images have minimal information and typically contain long series of pixels that have similar intensity values. An image whose pixels all have the same gray value will have zero entropy.

In the proposed entropy-based weighting scheme, we use the entropy calculation proposed in (Kapur et al. 1985). Given a gray value k, the *a posteriori* entropy of an image is defined by:

$$H'_n = -P_k \ln P_k - (1 - P_k) \ln(1 - P_k) \tag{9}$$

where

$$P_k = \sum_{i=0}^{k} p_i, \quad 1 - P_k = \sum_{i=k+1}^{L-1} p_i \tag{10}$$

If we maximize H'_n, we obtain a trivial result:

$$P_k = 1 - P_k = \frac{1}{2} \tag{11}$$

As proposed in (Kapur et al. 1985), this problem is solved by using an objective function:

$$\psi(k) = \ln P_k(1 - P_k) + \frac{H_k}{P_k} + \frac{H_n - H_k}{1 - P_k} \tag{12}$$

where

$$H_k = -\sum_{i=0}^{k} p_i \ln p_i, \quad H_n = -\sum_{i=0}^{L-1} p_i \ln p_i \tag{13}$$

The gray value k, which maximizes $\psi(k)$, can be used as the threshold value because it maximizes the distinction between the object and the background in the picture. However, if the distributions of the object and the background are uniform, then (12) is equivalent to

$$\psi(k) = \ln k(L - 1 - k) \tag{14}$$

and reaches the maximum when $k = \frac{1}{2}(L-1)$. This property is a drawback when being applied to product surface defect detection. It decreases the adaptability of the defect detection algorithm because the distributions of product surface images are typically uniform, as indicated in Fig. 3.

To overcome this drawback, we combine the entropy objective function $\psi(k)$ with Otsu's method, i.e., replace W by $\psi(k)$ in (7), to create a new objective function:

$$k^* = \underset{0 \le k \le L-1}{\arg\max} \{\psi(k)(\omega_1(k)\mu_1^2(k) + \omega_2(k)\mu_2^2(k))\} \tag{15}$$

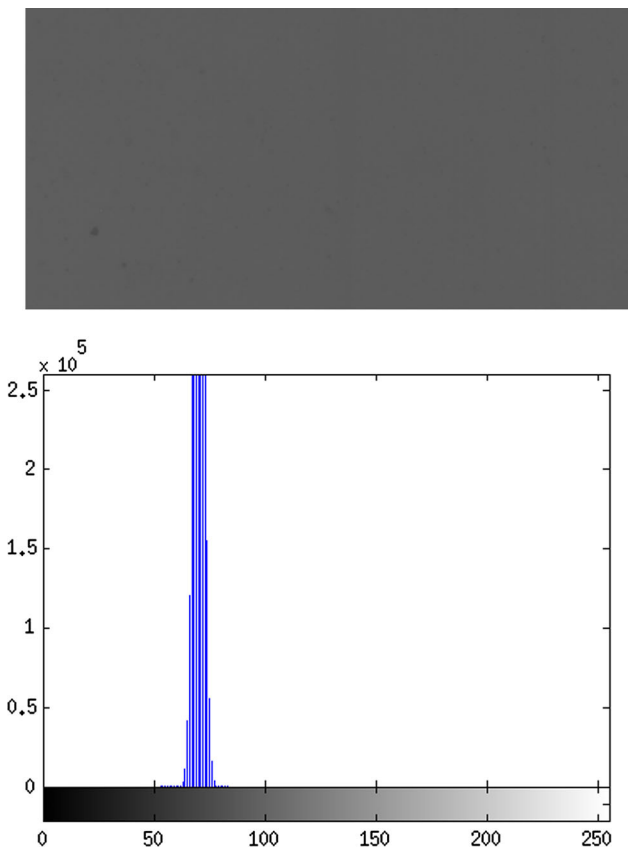

Fig. 3 Sample wafer surface (*top*) and its corresponding histogram (*bottom*)

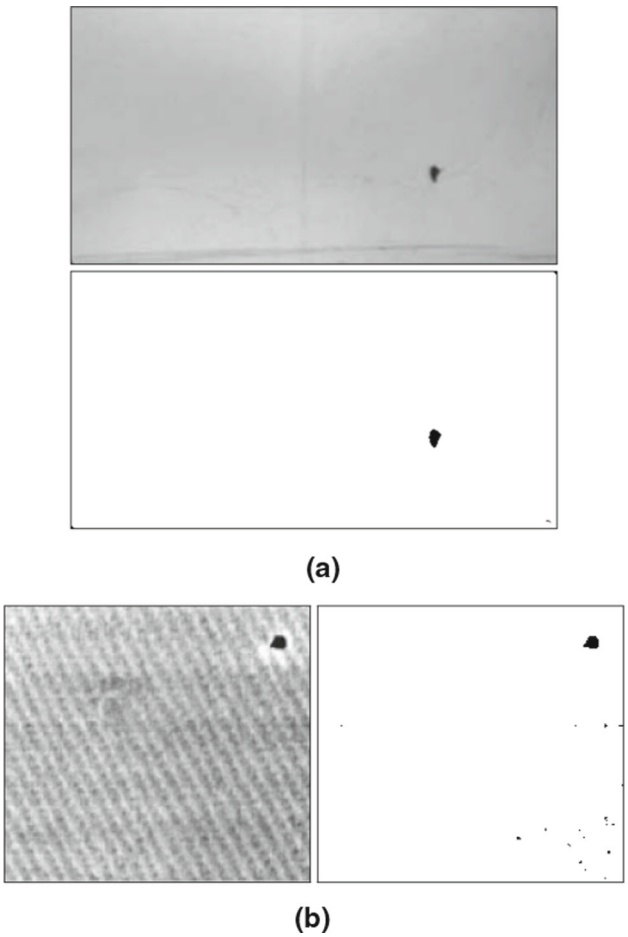

Fig. 4 Performance of proposed method for trivial cases. **a** Using sample from Fig. 1. **b** Using sample from Fig. 2

We obtain the value of k^* as the optimal value for threshold selection. The primary concept of the proposed weighting scheme for Otsu's method is simultaneously maximizing the image entropy and the between-class variance of the thresholded image. In visual inspection, we attempt to reveal as many defect regions as possible. Because defect regions are extremely small compared with the product surface area and appear to be random, the threshold that maximizes the image entropy also maximizes the number of revealed pixels from the defect regions.

4 Experimental results

In this section, we demonstrate the effectiveness of the proposed method in defect detection. Furthermore, we used our own implementations of Otsu's method and four weighting schemes from (Ng 2006; Fan and Lei 2012; Ng et al. 2013; Liu et al. 2014) for performance comparison. All six methods, including ours, were evaluated using a set of images. Then, we compared the thresholding results produced by the six methods. In methods that require assigning parameters, we used the recommended parameters given by their authors.

Figure 4 illustrates the performance of the proposed method in trivial cases of defect detection. The thresholding results produced by the proposed method were as effective as the results from the other methods, which are indicated in Figs. 1 and 2. The experiment also confirms that the proposed weighting scheme is robust when being applied in various types of product surface patterns.

Figure 5a displays the sample test image, which has several defect regions, for comparison of the different methods. Figure 5b–g shows the thresholding results from the six mentioned methods. We can observe that the output threshold value from the proposed method reveals other defect regions apart from the clearly visible regions, whereas the other methods fail or detect only the obvious region. Figure 6 displays the magnified region that is marked by a white rectangle in Fig. 5a. In this figure, the results from Otsu's method and the Gaussian-weight valley-emphasis method are not included because they are not viable for detecting defect regions in the sample test image.

For further evaluation, we used another set of ten same-sized images of wafer surfaces whose resolution was 3000×1500 pixels; the number of gray levels was $L = 256$. The images used in this experiment were acquired from an automated optical inspection system. In each image, there were several defect regions that were extremely small compared to the size of the image. The results are presented in Tables 1 and 2. As indicated in the tables, the number of defect regions detected by the proposed method is higher than that of the other methods. Furthermore, the output thresholds from the proposed method have the lowest deviation, whereas the output thresholds from the other methods vary significantly. This indicated the stability of our system when being applied to data produced by the same system.

5 Conclusion

Visual inspection is a commonly used nondestructive testing method because it does not require complex equipment; automatic thresholding is one of its promising approaches. Recently, several improved versions of Otsu's method, an automatic clustering-based thresholding technique, have been developed to enhance its performance in defect detection. In this research work, we proposed a novel visual inspection algorithm, an improvement to Otsu's method. Our proposal uses entropy theory and does not require manual parameter tuning. The algorithm is fully automatic and

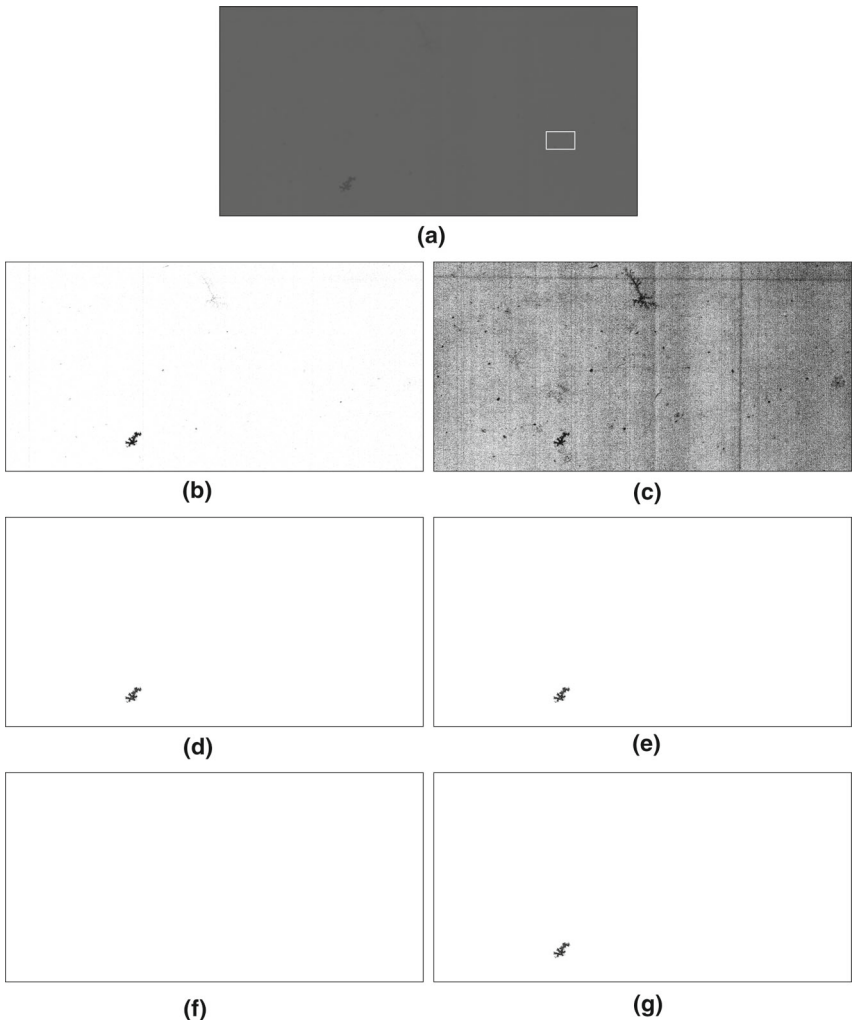

Fig. 5 Results of defect detection using various methods. **a** Sample test image. **b** Proposed method. **c** Otsu's method. **d** Valley-emphasis. **e** Neighborhood valley-emphasis, 11 neighbors. **f** Gaussian-weight valley-emphasis, $\sigma = 5$. **g** Fisher's LDA

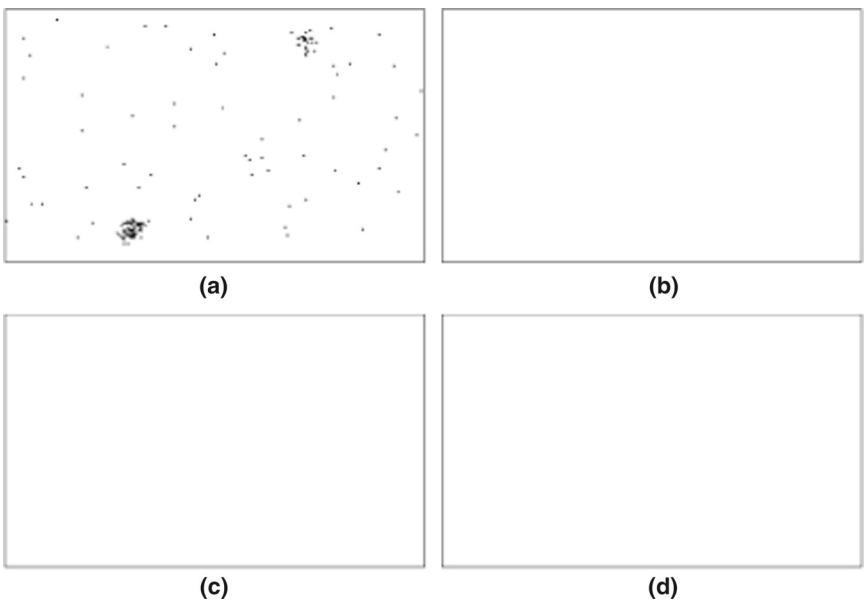

Fig. 6 Results of defect detection (magnified). **a** Proposed method. **b** Valley-emphasis. **c** Neighborhood valley-emphasis. **d** Fisher's LDA

Table 1 Pixel counts of defect regions

Image no.	Proposed method	Valley-emphasis	Neighbor VE	Fisher's LDA
1	1032	3	3	3
2	67	54	0	0
3	149	2	0	0
4	52	9	0	0
5	179	114	115	114
6	382	1	0	0
7	78	67	37	67
8	1021	662	629	662
9	39	3	3	3
10	15	1	0	0

Table 2 Output thresholds

Image no.	Proposed method	Valley-emphasis	Neighbor VE	Fisher's LDA
1	61	53	53	53
2	85	87	110	105
3	82	99	107	102
4	83	90	119	114
5	85	102	101	102
6	82	103	112	107
7	84	89	102	89
8	87	119	123	119
9	60	57	57	57
10	83	87	95	90
SD	9.97	20.25	24.10	22.44

capable of detecting extremely small defect regions compared with the product surface area. Moreover, because of the simplicity in calculation, the proposed method can be used for real-time processing and deployed in embedded systems that have limited computational resource. Several experiments were conducted comparing existing methods and the proposed method. In these experiments, the number of defect regions that the proposed method detected was greater than that of the other methods. Moreover, the output thresholds from the proposed method have the lowest deviation, implying the stability of the proposed method. The experimental results confirmed the efficiency of the proposed method over other techniques in surface defect detection. For future research, we will conduct a deeper examination of the entropy characteristics of visual data to enhance the performance of the entropy-based defect detection algorithm.

Acknowledgements This study was funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01057518).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Al-Tairi ZH, Rahmat RW, Saripan MI, Sulaiman PS (2014) Skin segmentation using YUV and RGB color spaces. J Inf Process Syst 100(2):283–299

Asha V, Bhajantri N, Nagabhushan P (2012) Automatic detection of texture-defects using texture-periodicity and Jensen-Shannon divergence. J Inf Process Syst 80(2):359–374

Bhajantri N, Kumar RP, Nagabhushan P (2013) Discriminatory projection of camouflaged texture through line masks. J Inf Process Syst 90(4):660–677

Chaki N, Shaikh SH, Saeed K (2014) A Comprehensive survey on image binarization techniques. In: Kacprzyk J (ed) Exploring image binarization techniques. Springer, New Delhi, pp 5–15

Fan JL, Lei B (2012) A modified valley-emphasis method for automatic thresholding. Pattern Recognit Lett 330(6):703–708

Gholizadeh S (2016) A review of non-destructive testing methods of composite materials. Proc Struct Integr 1:50–57

Hussain A, Abbasi AR, Afzulpurkar N (2012) Detecting & interpreting self-manipulating hand movements for student's affect prediction. Human-centric Comput Inf Sci 20(1):1–18

Kapur JN, Sahoo PK, Wong AKC (1985) A new method for gray-level picture thresholding using the entropy of the histogram. Comput Vis Graph Image Process 290(3):273–285

Liao PS, Chen TS, Chung PC (2001) A fast algorithm for multilevel thresholding. J Inf Sci Eng 170(5):713–727

Liu Z, Wang J, Zhao Q, Li C (2014) A fabric defect detection algorithm based on improved valley-emphasis method. Res J Appl Sci Eng Technol 70(12):2427–2431

Ng HF (2006) Automatic thresholding for defect detection. Pattern Recognit Lett 270(14):1644–1649

Ng HF, Jargalsaikhan D, Tsai H C, and Lin C Y (2013) An improved method for image thresholding based on the valley-emphasis method. In: Signal and information processing association annual summit and conference (APSIPA), 2013 Asia-Pacific, Kaohsiung, Taiwan. IEEE, pp 1–4

Otsu N (1979) A threshold selection method from gray-level histogram. IEEE Trans Syst Man Cybern 90(1):62–66

Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 130(1):146–168

Shankar N, Zhong Z (2005) Defect detection on semiconductor wafer surfaces. Microelectron Eng 770(3–4):337–346

Tolba AS, Raafat HM (2015) Multiscale image quality measures for defect detection in thin films. Int J Adv Manuf Technol 790(1):113–122

Uddin J, Islam R, Kim J-M (2014) Texture feature extraction techniques for fault diagnosis of induction motors. J Converg 50:15–20

Verma O, Jain V, Gumber R (2013) Simple fuzzy rule based edge detection. J Inf Process Syst 90(4):575–591

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com