Front. Electr. Electron. Eng. China 2011, 6(2): 328–338
DOI 10.1007/s11460-011-0139-x

RESEARCH ARTICLE

Alexander DROBCHENKO, Joni-Kristian KAMARAINEN, Lasse LENSU, Jarkko VARTIAINEN, Heikki KÄLVIÄINEN, Tuomas EEROLA

Thresholding-based detection of fine and sparse details

Abstract Fine and sparse details appear in many quality inspection applications requiring machine vision. Especially on flat surfaces, such as paper or board, the details can be made detectable by oblique illumination. In this study, a general definition of such details is given by defining sufficient statistical properties from histograms. The statistical model allows simulation of data and comparison of methods designed for detail detection. Based on the definition, utilization of the existing thresholding methods is shown to be well motivated. The comparison shows that minimum error thresholding outperforms the other standard methods. Finally, the results are successfully applied to a paper printability inspection application, and the IGT picking assessment, in which small surface defects must be detected. The provided method and measurement system prototype provide automated assessment with results comparable to manual expert evaluations in this laborious task.

Keywords adaptive thresholding, paper quality inspection, picking, machine vision

1 Introduction

The motivation for this work originates from the need for paper and board surface defect detection as part of their quality control. The defects are very small, but they can be manually spotted by a naked eye under oblique lighting (flat surface). Such tasks, however, are laborious and prone to errors, and therefore, their automation is under continuous investigation in the paper and printing industry. In this study, we present a system, including both machine vision hardware and algorithms, which solves the problem and reaches the required industrial precision. Since the problem can be solved by using a technique as simple as binary thresholding, we include a deeper analytical study of a more general class of surface defects, i.e., fine and sparse details, and their detection by thresholding. The purpose is to provide a general solution and analysis for similar quality inspection problems.

Thresholding-based segmentation can be found in the core of many machine-vision-based inspection applications, e.g., Refs. [1–3]. Several analytically sound and general solutions do exist, but in practice, thresholding is often based on application-specific heuristics. The use of heuristics usually prevents reusability, and the methods become sensitive to small changes in the problem setting. This study provides a more general solution to adaptively detect a small number of fine details from the background. The solution is designed by defining the appearance of the details analytically, and by studying the performance of well-known adaptive non-parametric thresholding methods.

Binary thresholding is one of the most common and essential operations in digital image processing. In several applications, thresholding is used at some point of the algorithm. Even though the operation itself is very simple, the problem of selecting the optimal threshold value is not trivial. For a single image, the optimal value can be selected manually, but there exist adaptive thresholding methods designed to automatically estimate the optimal threshold. Due to its importance, adaptive thresholding has been studied for a few decades, and a wide variety of different methods have been proposed (e.g., see Ref. [4]). The extensive work in the past should have resulted with a proper method for any problem. Since the different methods emphasize different properties, selecting the method for a specific case depends on the problem's characteristics.

Most methods perform well when the areas of the image's foreground and background are of sufficiently equal size, and the gray-level values have substantially

Received June 24, 2010; accepted January 26, 2011

Alexander DROBCHENKO, Joni-Kristian KAMARAINEN, Lasse LENSU, Jarkko VARTIAINEN, Heikki KÄLVIÄINEN, Tuomas EEROLA (✉)
Machine Vision and Pattern Recognition Laboratory, Department of Information Technology, Faculty of Technology Management, Lappeenranta University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland
E-mail: tuomas.eerola@lut.fi

non-overlapping distributions [4]. However, when either one of these assumptions is invalid, major difficulties can be encountered. This is the case in this study, where the motivation originates from a problem in which small paper surface defects must be automatically detected. An example of the proposed framework is related to the IGT picking assessment, which is a *de facto* standard for paper printability (runnability and print quality) evaluation used in the paper and printing industries. The assessment is based on test-printed paper samples. The samples can be imaged into a digital form where the low gray-level values correspond to the paper surface, and the high gray-level values correspond to small defects on the surface. The proportion of pixels related to the defects is typically very small, making the gray-level histograms almost unimodal. In addition, the gray-level values of both the defects and surface overlap significantly. The spatial distribution of the defects can be considered as random, i.e., the spatial relationships are meaningless. Therefore, it is motivating to apply global thresholding methods.

Successful methods do exist for detecting fine details, such as surface defects, if the underlying surface contains a structure, e.g., Ref. [5], or a training set for the defects is available, e.g., Ref. [6]. However, thresholding provides a general solution for the problem if the surface imaging setup is well designed, and a proper thresholding method is selected. This study assesses the ability of adaptive thresholding to emphasize fine and sparse details. The problem is solved by defining the necessary statistical properties in the histogram domain and by analyzing the performance of well-known and well-performing adaptive thresholding methods for general purposes and methods specifically applicable to unimodal histograms.

2 Fine and sparse details on noisy backgrounds

Methods to detect fine and sparse details, and their possible application areas are explained first. In the case of this study, the problem was to automatically perform visual assessment of IGT picking samples. An IGT printability tester prints a test pattern on a paper sample with increasing speed using tacky printing oils with known viscosity (standards related to the operation are ISO 3783, TAPPI T499, and SCAN-P 63). The proportion of visible defects, e.g., fiber buffing or coat tearing, is assumed to provide information about the printing properties of a particular paper type (printability including runnability and print quality). The longer is the distance from the print starting point to the beginning of fiber buffing, the higher is the maximum speed allowed in a real printing process.

The IGT printability tester is a mechanical device that does not produce any image data of the samples. The device consists of an oil applicator and a printing unit in which a printing cylinder produces a test pattern 200 mm long and 10 mm wide. The printing speed increases linearly toward the end of the print. To get the image data, the paper samples are digitally imaged under oblique lighting, see Fig. 1(a). The gray-level histogram of the strip can be seen in Fig. 1(b).

In Fig. 1(a), the defects are not clearly visible, and the image suffers from distortions characteristic to board strips (e.g., cockling). However, after a proper image enhancement, the defects appear as tiny spots having higher intensity than the surrounding noisy background, see Fig. 2.

Separation of the image data into background (paper surface) and foreground (defects) is more understandable in Fig. 2. For further processing, however, a suitable threshold value using which the defects on the paper surface begin to appear in the foreground must be selected. Based on a representing set of test samples, it was found that the proportion of pixels representing the defects was 0.1%–5.0% of the total image area, and they partly shared the intensity values with the background pixels. As a consequence, the background and foreground are mixed into a nearly unimodal gray-level histogram, making the selection of a threshold value a very difficult problem.

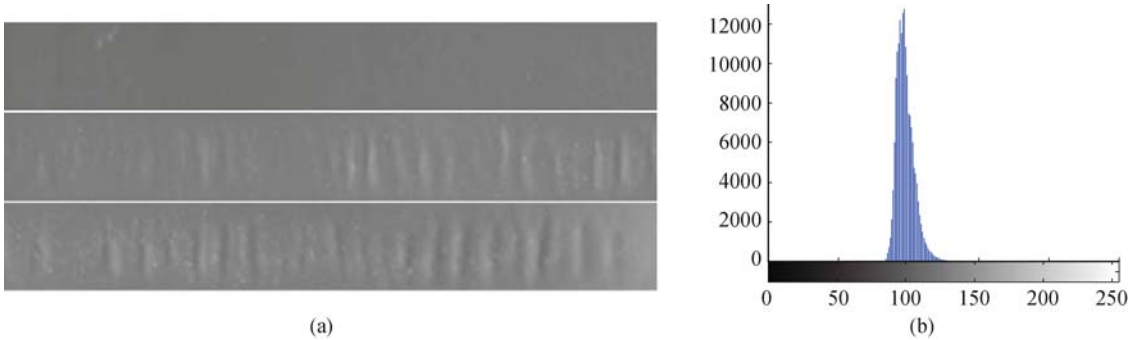

Fig. 1 (a) Image of an IGT picking sample (coated board) captured under oblique lighting. The whole strip is divided into three parts for a better presentation. The adjacent pieces from left to right (printing speed increases) are arranged from top to down. (b) Gray-level histogram of the image

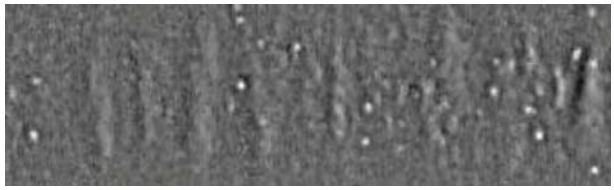

Fig. 2 Enhanced subimage of the IGT picking sample

The nature of the defects and their presence in the given problem have motivated us to introduce a concept of *fine and sparse defects*. It should be noted that the terms *details* and *defects* are used interchangeably in this study. The fine and sparse defects are small (fine) and isolated (sparse) signal patches comprising only a minority of the total image area. Their intensities are close to or mixed with the background intensity range. To study the problem more extensively, a statistical model of the defects must be derived. Based on the statistical model, the behavior of different thresholding techniques can be analytically studied since the effects of all model parameters can be studied.

2.1 Model for fine and sparse details

If the spatial relationships are neglected, image pixels can be considered as realizations of a random variable. For a sufficiently large image, the gray-level histogram corresponds to the probability density function of the random variable, and therefore, it is sufficient to model the probability density function (PDF) to model the details.

A noisy background can be modeled with a single probability density function, and the foreground can be treated as a mixture of probability density functions. Finally, the whole PDF consists of a weighted sum of the PDFs for both the foreground and background. It should be noted that if the both distributions are modeled with Gaussians, the solution is obvious, i.e., the minimum error thresholding, which will be described later. That model would, however, be too simple for more complicated situations, and therefore, we assume that the foreground is actually a sum of distributions.

Intensities of the background pixels can be modeled by a value of a random variable having the normal distribution $N(\mu_b, \sigma_b)$ with the mean μ_b and standard deviation σ_b. A single defect can be modeled by a low probability (low a priori) random variable that adheres to the normal distribution

$$P_d(i) = \frac{1}{\sqrt{2\pi}\sigma_d(i)} \mathrm{e}^{\frac{-(x-\mu_d(i))^2}{2\sigma_d(i)^2}}, \tag{1}$$

where $\mu_d(i)$ and $\sigma_d(i)$ denote the intensity mean and standard deviation for the ith defect, and $P_d(i)$ corresponds to the a priori probability of encountering the defect. However, since a single defect is highly localized (concentrated near to a single spatial location), $P_d(i)$ corresponds to a proportional spatial size of the defect rather than a true a priori probability. Correspondingly, the proportional spatial size of the background is

$$P_b = 1 - \sum_i P_d(i) \ . \tag{2}$$

Now, the resulting histogram of fine and sparse defects on a noisy background depends solely on the set of parameters $\{\mu_b,\ \sigma_b,\ \mu_d(i),\ \sigma_d(i),\ P_d(i)\}$. Finally, the composite probability density function that defines the expected shape of a histogram is

$$f(x) = P_b \frac{1}{\sqrt{2\pi}\sigma_b} \mathrm{e}^{\frac{-(x-\mu_b)^2}{2\sigma_b^2}} + \sum_i P_d(i) \frac{1}{\sqrt{2\pi}\sigma_d(i)} \mathrm{e}^{\frac{-(x-\mu_d(i))^2}{2\sigma_d(i)^2}}. \tag{3}$$

One more consideration is related to the distributions of $\mu_d(i)$ and $\sigma_d(i)$. The simplest class of sparse defects has a uniform distribution $\mu_d(i) \sim U(a, b)$, or the normal distribution $\mu_d(i) \sim N(\mu_{\mu_d}, \sigma_{\mu_d})$, and the standard deviation $\sigma_d(i) \sim N(\mu_{\sigma_d}, \sigma_{\sigma_d})$. It should be noted that the number of defects can also be obtained from a random distribution, but it affects only the foreground/background ratio and not the results.

2.1.1 Model visualization

For visualization purposes, a spatial-domain model that corresponds to the model in the gray-level histogram domain must be defined. This model includes a histogram and a mapping from the histogram to the spatial domain. The spatial model must be consistent with the histogram model and have the same parameters — the resulting artificial image must yield the same histogram as obtained via the histogram model.

First, the image background is generated using a random variable with the same distribution and parameters μ_b and σ_b, as described for the histogram model. Next, the defects are randomly seeded on the noisy background. For each defect, the area is derived in accordance with the total image size and the proportional defect size $P_d(i)$. Finally, the values at each defect area are derived from the corresponding random variable, $N(\mu_d(i), \sigma_d(i))$. To vary also the size, the proportional areas can be derived from $P_d(i) = N(\mu_{P_d}, \sigma_{P_d})$. It should be noted, however, that if a certain foreground/background ratio is required, the proportional size $P_d(i)$ must be normalized to achieve the requested ratio. An example of a generated image is shown in Fig. 3.

Fig. 3 Example image with fine and sparse details generated using the visualization of the histogram model ($\mu_b = 0.3$, $\sigma_b = 0.055$, $i = 1, 2, \ldots, 50$, $\mu_{\sigma_d} = 0.01$, $\sigma_{\sigma_d} = 0.002$, foreground/background ratio = 0.015, $\mu_{P_d} = 5$, $\sigma_{P_d} = 3$). The intensity values are scaled for visualization

3 Detail detection

The definition of details in Sect. 2 was carried out in the histogram domain, and thus, methods for histogram-based global thresholding can be used to separate the background and foreground. In this section, the methods are discussed, and their detection performance is studied using the defined statistical model.

3.1 Defect segmentation by thresholding

Thresholding methods are based on the assumption that the gray-levels of pixels belonging to objects (the foreground) are substantially different from the gray-levels of pixels belonging to the background. However, the characteristics of the histogram and the spatial distribution of the objects make some thresholding methods more suitable to a specific task than the others.

Based on the definition of the details, suitable thresholding methods were searched from relevant literature. First, the most popular and well-performing general-purpose thresholding methods were considered, and then, several methods designed especially for unimodal histograms were studied.

3.1.1 General gray-level thresholding methods

General thresholding methods should perform well when 1) the foreground objects and background constitute proportionally the same sizes in an image, and 2) the gray-level values of the objects and background possess substantially distant and non-overlapping distributions. When these constraints can be met, one of the most popular methods is Otsu's method [7]. Methods by Refs. [8] and [9] have been shown to outperform many others in comparisons [4]. Therefore, these methods are good candidates as general thresholding methods for the given problem. In the following, these three methods will be briefly reviewed.

Otsu's thresholding method The method is based on the idea of finding a threshold value that minimizes the within-class variance of the resulting foreground and background classes [7]. The optimal threshold T is calculated by minimizing the criterion function

$$J(T) = \frac{P_1(T)\sigma_1^2(T) + P_2(T)\sigma_2^2(T)}{\sigma^2}, \tag{4}$$

where σ^2 is the gray-level total variance estimated from the histogram.

Otsu's thresholding is one of the most widely used and cited threshold estimation method. The method is very robust, and it provides thresholding results of considerable quality in a vast variety of cases. However, the method usually fails to find an appropriate threshold value if the number of foreground pixels is less than 5% of the total image size.

Kapur's thresholding method A method based on entropy has been proposed by Kapur et al. [9]. The method maximizes class entropies, which can be interpreted as measures of class compactness and separability. In this case, the criterion function can be given as

$$\begin{aligned} J(T) = &- \sum_{g=0}^{T} \frac{h(g)}{P_1(T)} \log \frac{h(g)}{P_1(T)} \\ &- \sum_{g=T+1}^{n} \frac{h(g)}{P_2(T)} \log \frac{h(g)}{P_2(T)}, \end{aligned} \tag{5}$$

where n is the maximum length of a histogram. $J(T)$ is maximized to obtain maximum information between the object and background distributions in the image. The discrete value T, which maximizes $J(T)$, is the threshold.

Kittler's thresholding method Kittler and Illingworth have proposed an algorithm whose cost function to be optimized is based on the Bayesian classification rule [8]. In this method, it is assumed that the components of a bi-modal histogram in a gray-level image $h(g)$ are normally distributed. Normal distributions are defined by their means μ_i, standard deviations σ_i, and a priori probabilities P_i. For a case of two different classes ($i = 1, 2$), the background and foreground, and given a threshold T, the parameters can be estimated from the following:

$$\begin{aligned} P_i(T) &= \sum_{g=a}^{b} h(g), \\ \mu_i(T) &= \frac{1}{P_i(T)} \sum_{g=a}^{b} h(g)g, \\ \sigma_i^2(T) &= \frac{1}{P_i(T)} \sum_{g=a}^{b} (g - \mu_i(T))^2 h(g), \end{aligned} \tag{6}$$

where

$$a = \begin{cases} 0, & i = 1, \\ T+1, & i = 2; \end{cases} \text{ and } b = \begin{cases} T, & i = 1, \\ n, & i = 2. \end{cases} \tag{7}$$

Now, the criterion function can be calculated as

$$J(T) = 1 + 2[P_1(T)\log\sigma_1(T) + P_2(T)\log\sigma_2(T)] - 2[P_1(T)\log P_1(T) + P_2(T)\log P_2(T)]\ , \quad (8)$$

and the minimum-error threshold can be computed by minimizing the criterion $J(T)$. It should be noted that the optimum for the criteria in Eqs. (4), (5) and (8) can be efficiently found by searching over all possible values of the threshold T.

3.1.2 Unimodal histogram thresholding methods

As already discussed, most thresholding methods work in the general case of bi- or multi-modal gray-level histograms. In such cases, all modes in the histogram are considered to represent different objects, or the background. However, the properties of the fine details, as seen above, cause the images to have a close-to-unimodal distribution. Therefore, the information available for distinguishing the details is hidden in something that resembles normal background noise. Fortunately, this is not a new problem to the thresholding field, and several different methods have been proposed to resolve the problem. Some of these methods have a very sound basis, while the others are more or less ad-hoc solutions. Two unimodal thresholding methods from the literature, Tsai's method [10] and Rosin's method [11], are interesting because of their adaptive nature. These two should be applicable to the given problem.

Tsai's thresholding method Tsai has introduced two similar approaches to image thresholding using smoothed histograms, one of which is especially intended for unimodal histograms [10].

The first approach searches peaks and valleys in the histogram smoothed with a Gaussian kernel. The smoothing level is adjusted to make the histogram to contain exactly the same number of peaks as the desired number of thresholding levels. The valleys between the peaks are selected as the threshold values. In a case where the number of peaks is less than the desired number after using the smallest possible Gaussian kernel for smoothing, additional threshold values are selected based on the curvature maxima of the histogram.

The second approach utilizing curvature is especially intended for unimodal histograms, and represents a custom case of the first approach. In a case where only one peak can be found in the histogram, which is true in the unimodal case, the threshold value is selected as the intensity value at which the histogram reaches its maximum curvature. The curvature values are calculated from [10]

$$K_t = \frac{1}{R}\sum_{j=1}^{R}|\psi_{t+j} - \psi_{t-j}|, \quad (9)$$

where

$$\psi_t = \frac{1}{R}\sum_{j=1}^{R}\frac{h_k(t+j) - h_k(t-j)}{2j}, \quad (10)$$
$$h_k(i), \quad i = 1, 2, \ldots, R,$$

are the smoothed histogram values. This definition actually differs from the formula for an arbitrary curve curvature.

Rosin's thresholding method This is another method for binary thresholding in the case of unimodal histograms [11]. The algorithm for finding the threshold value is extremely simple:

Algorithm 1 Rosin's thresholding

1) A line is drawn from the maximum of the histogram to the last non-zero element of the histogram:

$$(\arg\max_i\ H_i, \max_i\ H_i) \quad \rightarrow \quad (\arg\max_i\left[H_i = 0 \text{ and } H_{i-1} \neq 0\right], 0),$$

where H_i is the ith element of the histogram.

2) The optimal threshold value is selected as the intensity value that maximizes the perpendicular distance between the line and the histogram.

This method definitely lacks intuitive motivation. The theoretical mathematical analysis shows that the method is almost insensitive to foreground pixels, and it actually determines the threshold value using information about the dominating background only [11].

3.2 Evaluation of thresholding methods

The constructed model allows detailed analysis of the different methods as functions of the model parameters since the distributions of both foreground and background are known. Selecting the optimal method for a specific application can be achieved by resolving which model parameters correspond to the variation in the acquired data.

In the evaluation of different adaptive thresholding methods, the most important success factors are 1) the number of undetected foreground pixels (false negatives), and 2) the number of falsely detected background pixels (false positives). The values, however, are dependent, and thus, should be combined to unambiguously compare the success between different methods. For the comparison, we adopt the enhanced Venkatesh-Kitchen discrepancy measure introduced in Ref. [12]. The discrepancy measure is originally designed for comparing edge detection methods, but it can be utilized here by setting the measure parameter values to $\alpha = 0.5$, $\beta = 0.5$, $\gamma = 0.0$, and $\delta = 0.0$. $\alpha = \beta = 0.5$ yields from the assumption that both the false positives and false

negatives are considered as equally erroneous mistakes. γ and δ are set to zero since they affect the spatially-dependent error factors, that is, errors near the correct values are considered less erroneous when compared to errors further away. This is justifiable in the case of edge detection, but for the fine and sparse details, the errors are equally erroneous whatever their location is.

In the IGT picking data, e.g., the statistical properties consistently change along the strips. In this specific case, the change can be modeled as increasing the a priori probability of encountering a defect $P_d(i)$ as demonstrated in Fig. 4. In Fig. 4, the histograms, computed from areas of the same size at different locations of the sample, have the same shape, but their amplitude increases.

Since the foreground/background ratio is clearly an important factor, it is motivating to inspect the performance of different methods as a function of this ratio. The behavior remains the same with different selections of parameters, but here the values are derived to correspond to the ones occurring in the IGT picking samples. The actual foreground/background ratios vary between 0.1%–5.0%, while the other characteristics correspond to the model parameters $\mu_b = 0.30$, $\sigma_b = 0.055$, $i = 1, 2, \ldots, 50$, $\mu_d(i) \sim U(0.50, 0.80)$, $\sigma_d(i) = N(\mu_{\sigma_d}, \sigma_{\sigma_d}) = N(0.01, 0.002)$. The behavior was inspected for ratios between 0.001 and 0.050, and the histograms and images corresponding to the two extreme values are shown in Fig. 5.

Surprisingly, a change in the a priori value does not

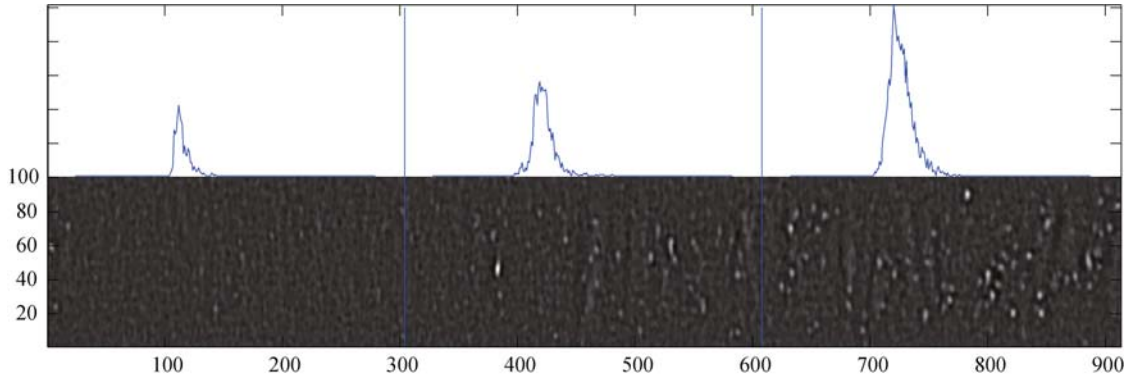

Fig. 4 Histograms of manually marked defects

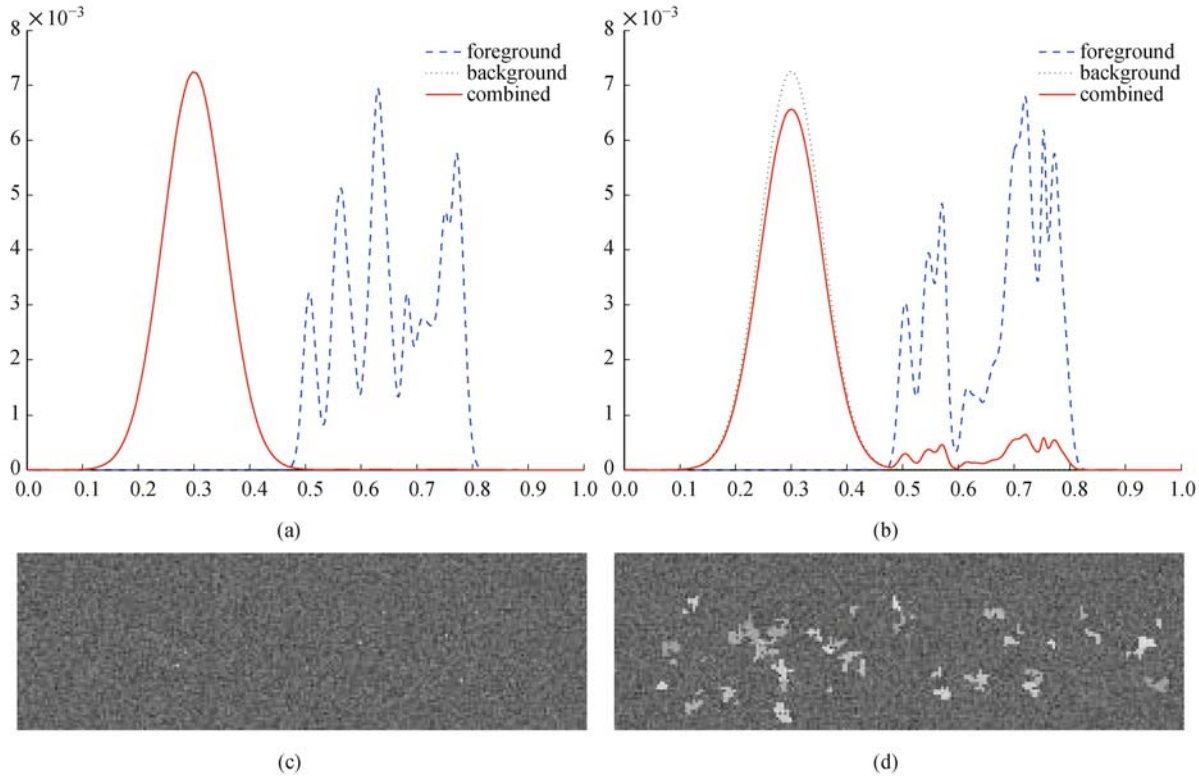

Fig. 5 Model-generated histograms and the corresponding artificial images. (a),(c) Foreground/background ratio ($\sum P_d(i)/P_b$) 0.001; (b),(d) 0.05. Note that the background is almost equal to the combined curve in subfigure (a) since the ratio is set to very low (nearly unimodal)

induce any significant change in the performance of the methods, as shown in Fig. 6. This behavior is natural for the methods based on the assumption of a bimodal distribution (Kittler). Otsu's method completely fails by detecting most of the background as foreground (Fig. 6(b)), and Kapur's method fails by missing a significant number of the foreground defects (Fig. 6(a)). These two general thresholding methods seem to be inapplicable to the given problem domain. Tsai's method performs well for a small number of defects, but it becomes unstable when the foreground/background ratio approaches 0.05, see Figs. 6(a) and 6(b). The success of the different methods becomes even more clear in the discrepancy graph in Fig. 6(c). The most successful methods are Kittler's and Rosin's methods.

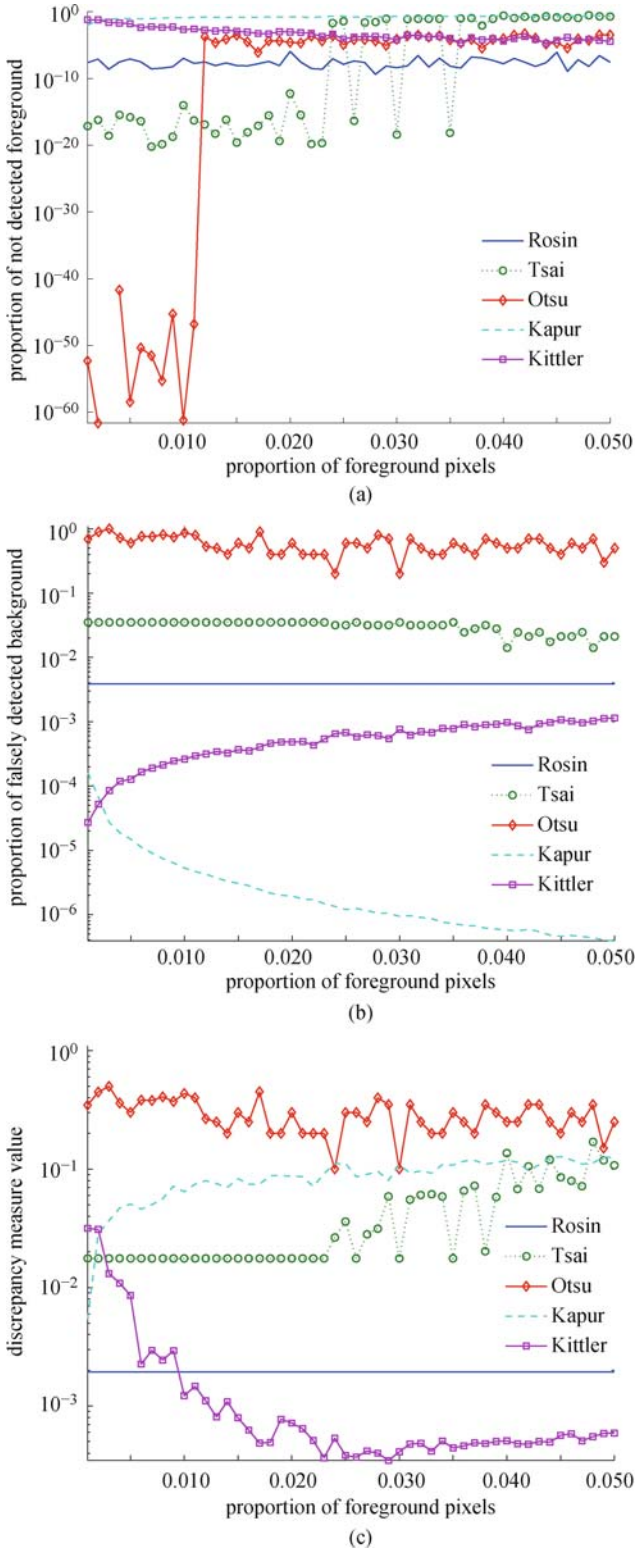

Fig. 6 Detection results for artificial data. (a) Proportion of not detected foreground pixels; (b) proportion of falsely detected background pixels; (c) discrepancy

The IGT data supports also the next study since the distribution of the paper defects can be modeled as a normal distribution, see Figs. 4 and 7.

Since the histogram in Fig. 7 posses a clearly unimodal distribution, it is motivating to study how the methods perform as a function of the level of unimodality. The sensitivity can be tested by varying the mean of the defect distribution, i.e., the mean of foreground distribution μ_{μ_d} was varied between 0.36 and 0.56, while the other parameters were kept constant. From the results in Fig. 8, it is apparent that Rosin's, Tsai's, and Kittler's methods perform equally well in the detection of defects, and only Otsu's method provides significantly worse results. The results are again most evident in the discrepancy graph in Fig. 8(c) (note that Kapur's method is omitted).

4 Experiments

In this section, results of the experiments with real IGT picking samples are reported. First, some example images of processed samples are shown, and then, a comparison of results by human experts and the implemented automatic system prototype is reported. The

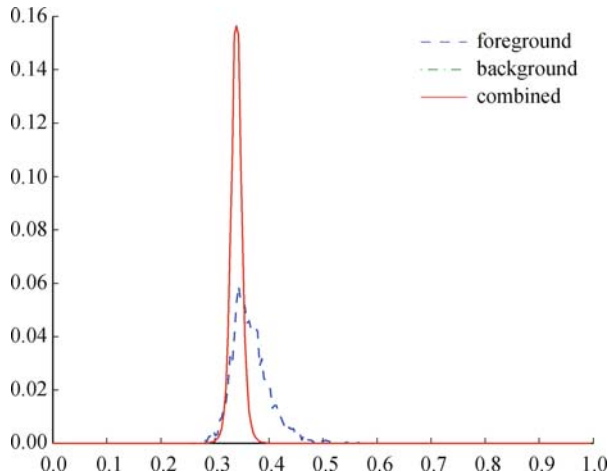

Fig. 7 Histograms of manually marked IGT picking defects and the background. The foreground/background ratio is $\sum_i P_d(i) = 0.005$, the background mean is $\mu_b = 0.345$, and the standard deviation $\sigma_b = 0.011$. The foreground mean is $\mu_{\mu_d} = 0.372$, and the standard deviation is $\sigma_{\sigma_d} = 0.037$. Note that the background is almost equal to the combined curve since the ratio is very low (nearly unimodal)

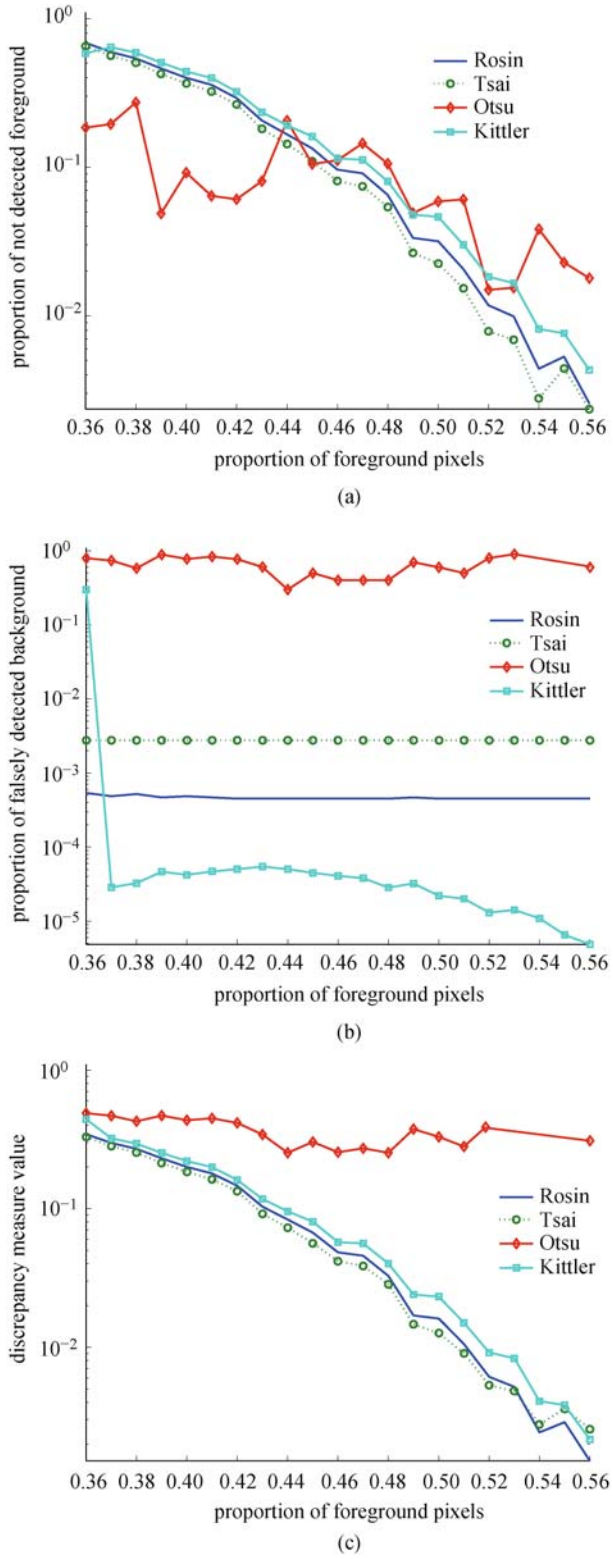

Fig. 8 Detection results for artificial data as functions of defect distribution mean μ_{μ_d}. (a) Proportion of foreground pixels that were undetected; (b) proportion of pixels that were falsely detected as background pixels; (c) discrepancy

prototype system consists of a Sony Cast Pro II desktop robot with adjustable light sources illuminating the sample at an oblique angle (see Fig. 9). The system is fully automatic after the sample to be measured has been placed on the robot's work table. The automatic

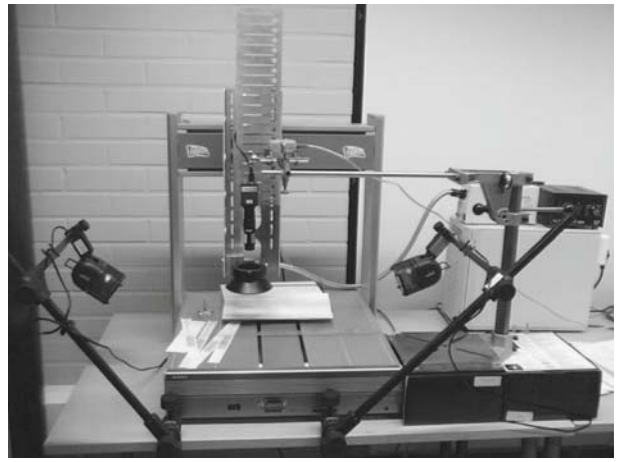

Fig. 9 Automatic IGT picking measurement system prototype: Sony Cast Pro II desktop robot, Allied Vision Technologies firewire camera (supported in Linux), Moritex micro objective, and standard halogen light sources

measurements were compared to two independent evaluations by human experts.

4.1 IGT picking images

Results of the proposed methods are demonstrated using images from IGT picking samples captured under oblique illumination.

4.1.1 Image enhancement

In common situations, images are or have been derived from images with only 8 bits per pixel, i.e., 256 intensity levels. Therefore, there is little motivation to generate histograms with more than 256 bins. However, certain image transformations, such as convolution filtering with floating point data processing, can dramatically increase the number of intensity levels and produce larger histograms, which allow more precise threshold estimation.

It is clear that certain image enhancement techniques may lead to more accurate thresholding results since sub-level accuracy is achieved. With fine and sparse details, it is beneficial to filter the image with a spot detection filter. By filtering with a simple $n \times n$ filter, the number of distinct intensities in an image can increase up to the factor of n^2 times. Such image preprocessing before constructing the histogram agrees with the results presented in Ref. [13]. In the case of IGT picking images, the following spot detection filter was used:

$$f = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & 1.5 & 3 & 1.5 & 0 & -1 \\ -1 & 0 & 3 & 6 & 3 & 0 & -1 \\ -1 & 0 & 1.5 & 3 & 1.5 & 0 & -1 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}. \quad (11)$$

The main reason for enhancing the IGT picking images was to remove undesired imaging effects related to the paper samples, such as cockling, the wavy appearance of the sample visible in the top row of Fig. 1(a). The thresholding results without the enhancement are demonstrated in Fig. 10, where all methods falsely detected the cockling instead of the defects.

4.1.2 Image examples

In Figs. 11 and 12, the images of four IGT picking samples, their enhanced versions, and the results for the five thresholding methods are shown. The results with the real images correspond to the results with the artificial data: two of the most promising methods are Rosin's method and Kittler's method. Rosin's method detected the more foreground pixels, but it also falsely detected a large number of background pixels. This is evident in Figs. 6, 11, and 12. Therefore, Kittler's method corresponds more precisely to what needs to be detected from the original images.

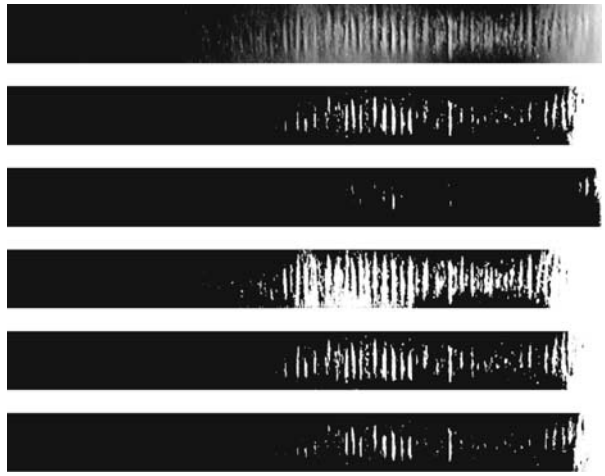

Fig. 10 Thresholding results for the image in Fig. 11(a) without the enhancement. Images from top to bottom are the aligned image, the results from the method by Rosin, Tsai, Otsu, Kapur, and Kittler

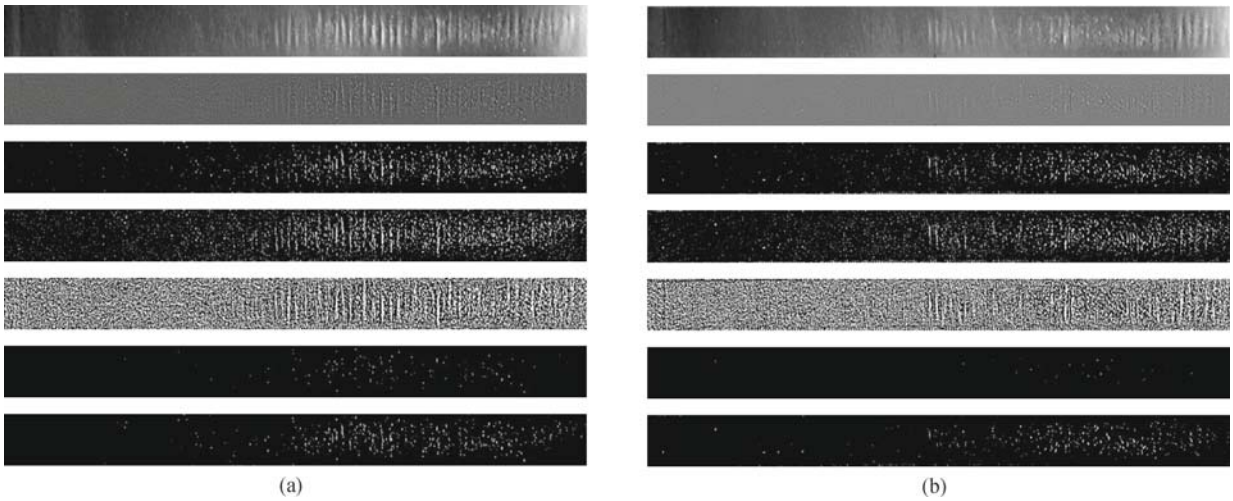

Fig. 11 Examples 1 and 2 of IGT picking images with the thresholding results. Images from top to bottom are the aligned image, the enhanced image, the results from the method by Rosin, Tsai, Otsu, Kapur, and Kittler

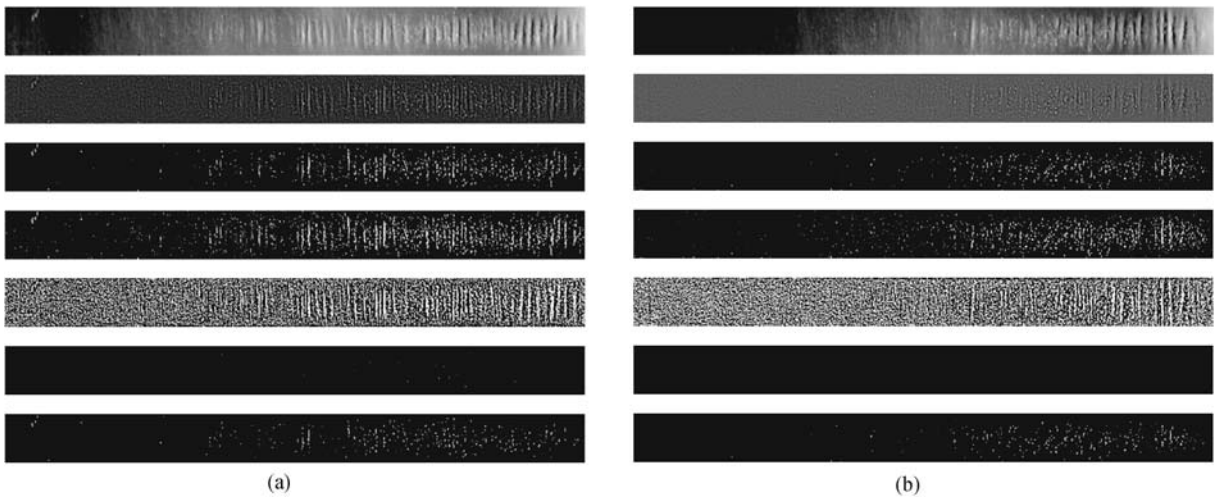

Fig. 12 Examples 3 and 4 of IGT picking images with the thresholding results. Images from top to bottom are the aligned image, the enhanced image, the results from the method by Rosin, Tsai, Otsu, Kapur, and Kittler

4.2 Automatic system vs human experts

The original motivation of this research was to design an automatic system comparable to the current practice, where the IGT samples are evaluated by human experts. In this section, a quantitative comparison between the automatic system and experts is provided.

For this experiment related to the end-use of the proposed method, five different paper types (P1, P2, P3, P4, and P5) and five random samples of each type were selected. The samples were first evaluated by two independent human experts, and then measured using the proposed system. The experts measured the distance from the beginning of the test print to the location where the picking was considered to start, i.e., the first place where severe defects occur. The same distance was measured by the system. It is important to note that deciding the starting location is often quite subjective. Based on the IGT protocol, the longer the distance from the beginning of the test print to the appearance of picking, the better the paper printability. The correlation coefficients from the two independent trials and the automatic measurement are given in Table 1. This experiment revealed the reason why automatic methods are investigated in the paper and printing industry: the expert-based evaluations are very subjective by their nature (see the paper type 5 in the table). On the average, the proposed system is well in accordance with the expert number 1 (E1) under the variation between the experts, and it also corresponds well to the opinion of the expert E2 except for the paper type 5.

Table 1 Correlation coefficients between the evaluators (E1 and E2) and the system (S)

	P1	P2	P3	P4	P5
E1 vs E2	0.93	0.96	0.70	0.96	0.37
S vs E1	0.70	0.98	0.76	0.57	0.77
S vs E2	0.73	0.92	0.50	0.62	0.01

5 Conclusions

In this study, a model for images with fine and sparse details was given based on the statistical properties of the image gray-level histograms. Based on the proposed model, several well-known and widely used thresholding methods were studied to evaluate their performance as functions of the model's variables. The most promising methods were proposed for use in the detection of fine and sparse details.

The proposed model aims to explain characteristics of real images containing different types of defects on a noisy background. An application introduced in this study is the automatic evaluation of IGT picking samples. Visualization of the model was shown to correspond to the real data, and the results with real images verified the analytical results. Furthermore, a full machine vision system was implemented, and the experiments revealed that the system could replace the subjective measurements by laboratory experts providing good precision and repeatable measurement results. Based on the study and the conducted experiments, Kittler and Illingworth's minimum error thresholding was selected as the most suitable method for the given task.

Acknowledgements The work was partially supported by the following projects: PapVision (http://www.it.lut.fi/project/papvision/) financed by the European Union (Tekes project No. 70049/03 and 70056/04), and the Academy of Finland (Project 204708).

References

1. De La Escalera A, Moreno L E, Salichs M A, Armingol J M. Road traffic sign detection and classification. IEEE Transactions on Industrial Electronics, 1997, 44(6): 848–859
2. Lia Q, Wang M, Gu W. Computer vision based system for apple surface defect detection. Computers and Electronics in Agriculture, 2002, 36(2–3): 215–223
3. Medina-Carnicer R, Madrid-Cuevas F J, Fernández-García N L, Carmona-Poyato A. Evaluation of global thresholding techniques in non-contextual edge detection. Pattern Recognition Letters, 2005, 26(10): 1423–1434
4. Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronical Imaging, 2004, 13(1): 146–165
5. Chan C H, Pang G K H. Fabric defect detection by Fourier analysis. IEEE Transactions on Industry Applications, 2000, 36(5): 1267–1276
6. Iivarinen J, Rauhamaa J, Visa A. Unsupervised segmentation of surface defects. In: Proceedings of the 13th International Conference on Pattern Recognition. 1996, 356–360
7. Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62–66
8. Kittler J, Illingworth J. Minimum error thresholding. Pattern Recognition, 1986, 19(1): 41–47
9. Kapur J N, Sahoo P K, Wong A K C. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing, 1985, 29(3): 273–285
10. Tsai D M. A fast thresholding selection procedure for multimodal and unimodal histograms. Pattern Recognition Letters, 1995, 16(6): 653–666
11. Rosin P L. Unimodal thresholding. Pattern Recognition, 2001, 34(11): 2083–2096
12. Fernández-García N L, Medina-Carnicer R, Carmona-Poyato A, Madris-Cuevas F J, Prieto-Villegas M. Characterization of empirical discrepancy evaluation measures. Pattern Recognition Letters, 2004, 25(5): 35–47

13. Baradez M O, McGuckin C P, Forraz N, Pettengell R, Hoppe A. Robust and automated unimodal histogram thresholding and potential applications. Pattern Recognition, 2004, 37(6): 1131–1148

Alexander DROBCHENKO received his M.Sc. degree in information processing (Comp. Sc.), from Lappeenranta University of Technology in 2004.

Joni-Kristian KAMARAINEN received MSc (Eng.) and DSc (Tech.) degrees, both in information processing (Comp. Sc.), from Lappeenranta University of Technology in 1999 and 2003, respectively. He is a founding member and vice director of the Machine Vision and Pattern Recognition Laboratory, Lappeenranta University of Technology where he was appointed Professor of Information Society Technologies in 2008. His research interests include computer vision, image analysis and pattern recognition. He is a chairman of Pattern Recognition Society of Finland, and member of International Association for Pattern Recognition (IAPR).

Lasse LENSU is an associate professor of computer science at Lappeenranta University of Technology (LUT), Finland. He received his M.Sc. (Tech.) degree in data communications in 1991, and Lic.Sc. and D.Sc. (Tech.) degrees in computer science in 2001 and 2002 from the Department of Information Technology of LUT. He received the title of docent in 2009. His research interests include biomolecular computing, especially applications of photoactive proteins, computational vision, and digital imaging and image processing applications. Lensu has been involved also in the information processing industry, technology transfer between academic institutions and companies, and several research projects.

Jarkko VARTIAINEN received his M.Sc. degree and Ph.D. degree (Doctor of Technology) in information processing (Comp. Sc.) in 2003 and 2007, respectively, from Department of Information Technology in Lappeenranta University of Technology, Lappeenranta, Finland.

Heikki KÄLVIÄINEN received his M.Sc. degree and Ph.D. degree (Doctor of Technology) in Computer Science in 1989 and 1994, respectively, from Department of Information Technology in Lappeenranta University of Technology (LUT), Lappeenranta, Finland. Since 1996 Kälviäinen has been a Professor of Computer Science. Currently he is a Head of Department of Information Technology, and a Head of Machine Vision and Pattern Recognition Laboratory. His primary research interests include machine vision, pattern recognition, and image processing and image analysis. Prof. Kälviäinen belongs to the governing board of IAPR, and he is a member of ACM, IEEE, and SPIE.

Tuomas EEROLA received his M.Sc. degree and Ph.D. degree (Doctor of Technology) in information processing (Comp. Sc.) in 2006 and 2010, respectively, from Department of Information Technology in Lappeenranta University of Technology, Lappeenranta, Finland. He is currently a postdoctoral researcher with Machine Vision and Pattern Recognition Laboratory, Lappeenranta University of Technology. His research interests include digital image processing, pattern recognition and image quality assessment.