Melnyk R.,
Kvit R.

MEASUREMENT OF MATERIAL SURFACE DEFECT INTENSITY BY DISTRIBUTED CUMULATIVE HISTOGRAM AND CLUSTERING

The object of research is a distributed cumulative histogram of a digital image and its advantages for automated determination of the location and intensity of defects of different nature on the surfaces of materials: metal, paper, etc. The technique considered in the study is aimed at minimization of human interference in the process of material surface control from the moment of its photographing to the moment of making a decision about the surface quality.

Three-dimensional distributed cumulative histogram (DCH) is presented as a two-dimensional image in which the pixel intensity corresponds to the third dimension – the number of pixels of a certain intensity in the original surface image. Informative distributed cumulative histogram (IDCH) is used to recognize black, dark and light defects, and to measure their intensity and location by the clustering algorithm. The average value of the pixel intensity in the columns and rows of the pixel matrix of the cumulative histogram image is calculated to estimate the intensity of the defects. Measurement of the intensity of defects is carried out in two ways: directly on the image of the surface sample and by comparing the image of the sample and the reference image of the sample without defects. To solve the problem, an algorithm of hierarchical clustering of data to rectangular segments of the surface image is used. In the image, each cluster is marked with a corresponding color of gray. The image for analysis is transformed using segmentation and inversion algorithms. This allows to get more accurate estimates of the intensity of light and dark defects. The clustering algorithm groups the image segments of the surface samples, as well as the images of the distributed cumulative histogram to group the level of surface damage. Distributed cumulative histogram was used to detect defects on the surface of materials as a method of linking the number and intensity of pixels to image coordinates. Cluster analysis helps to find their coordinates and intensity.

In comparison with known approaches, the method has a linear algorithmic complexity to the number of pixels in the input image, which allows to do a significant number of experiments to identify the types of surfaces of materials for use and the features of algorithms.

Keywords: *distributed cumulative histogram, clustering algorithm, hierarchical tree, histogram image, examples of defects.*

1. Introduction

The main task of an optical and computer control systems is to distinguish a defect from the background (brightness, edges, pattern, etc.) and apply a decision rule to determine which items should be classified as defects or a robust surface. Wide varieties of approaches of defects determination differ between themselves by features and extraction algorithms. For example, the paper [1] considers the probability of detecting varying size and magnitude defects in addition to the probability of false alarms and proposes an adaptive generalized likelihood ratio (AGLR) technique. The algorithm in [2] calculates the difference between the original signal and a smooth one in the log amplitude spectrum, and then a saliency map is obtained by transforming the difference into a spatial domain.

The unified approach for defect detection [3] consists of two phases: global estimation and local refinement. First, it roughly estimates defects by applying a spectral-based approach in a global manner. Then it locally refines the estimated region based on the distributions of pixel intensities. The paper [4] presents an automatic system based on Hough Transform, Principal Component Analysis and Artificial Neural Networks to classify three defects with well-defined geometric shapes: welding, clamp and identification hole.

The paper [5] describes the algorithm that extracts local statistical features from grey-level texture images decomposed with wavelet frames into sub bands of various orientations.

Some of the papers propose templates and neural networks to classify defects. For example, in the paper [6] orthogonal projection was applied to locate each pixel in a test image. Then, an image comparison was performed to mark

similar blocks in the test image. The block that best resembled the template was chosen as the new template. A predefined template-based image processing system is proposed to automatically detect of PCB soldering defects in the paper [7]. The proposed system consists of a scaled inspection structure, a camera, the algorithm of image processing merged with Fuzzy and template guided inspection process. In paper [8] Gaussian Mixture model is trained with the features extracted from a handful of defect-free texture samples. In a second step, the detection of texture samples is performed by comparing each pixel to the likelihood obtained in the training process. The paper [9] proposes artificial neural network based methodology to classify the defects.

In addition, many papers present defect using clustering. For example, detection methods based on the image segmentation the algorithm in paper [10] uses k-means algorithm to split the original image into regions based on Euclidean color distance to produce an over-segmentation result. In paper [11] cluster analysis (TCA), a method for automatic defect detection is based on three-dimensional image segmentation. Many fuzzy, C-Means, K-Means clustering methods are described in the papers [12–15]. These approaches are widely used for image segmentation, pattern recognition, finding the optimal segmentation threshold, classification and defect detection. For researchers the task is to choose simply, effective algorithm easily to be modified in new conditions.

Some of the above-mentioned approaches are quite complicated and time-consuming. Therefore, the development of a cluster algorithm for calculating the image of informative distributed cumulative histogram is a new technique having linear algorithmic complexity and is an actual task. It will allows to hold many experiments for defect detection in different material surfaces.

The object of research is a distributed cumulative histogram of a digital image and its advantages for automated determination of the location and intensity of defects of different nature on the surfaces of materials: metal, paper, etc. *The aim of research* is to minimize human intervention in the process of controlling the surface of the material from the moment of its photographing to the moment of making a decision on the quality of the surface.

2. Methods of research

To get accurate dimension and intensity data reflecting the image of the surface defects let's propose a new approach based on the concept of the cumulative histogram namely a distributed cumulative histogram (DCH). Dividing the image into categories of rows and columns let's increase the role of every pixel in the resulting picture reflecting a distribution of pixel frequencies in the spaces of intensity and one image coordinate: by *OX* or by *OY*. Therefore, let's distinguish two types of the DCH: a view from the image front side (*OX*) and a view from the left image side (*OY*).

To illustrate the base concept of the approach let's consider the image (250×250) of metal defects in Fig. 1, *a*. This and other examples for experiments and investigations were taken from resources [6, 16]. From this image let's copy two lines: one having defects and the other without them (Fig. 1, *b*). Ideally, these lines should have a height of one pixel. However practically it is difficult to cut out such a line and besides it is almost invisible as an image for users. Therefore, this lines have a height of eight pixels.
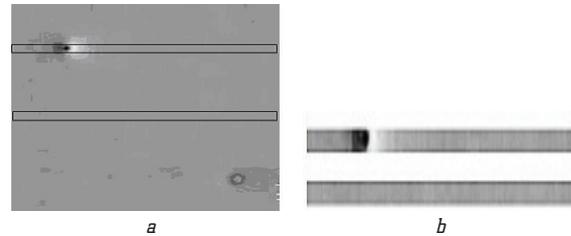
Fig. 1. Images for analysis: *a* – metal sample with defects; *b* – cropped lines with and without defects

Then let's calculate normalized cumulative histograms for all three samples (Fig. 2).
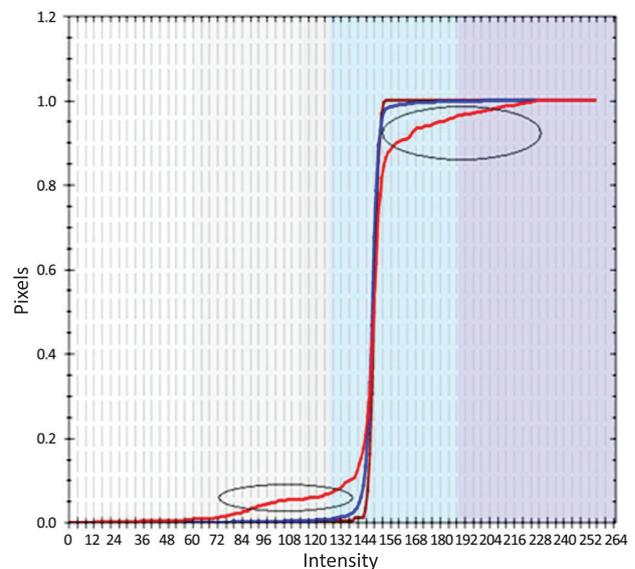
Fig. 2. Cumulative histograms of three metal images: sample and two small parts of it

It is possible to see that histograms of the full metal sample and the line without defects are close. The histogram of the line with defects differs essentially from previous ones particularly before and after a section responsible for a background color.

The reason for this fact is that a function sensitivity to the pixel intensity in one row is much greater than a function sensitivity for full image: $1/250 \gg 1/(250^{\wedge}250)$.

This fact concerns a mean intensity, cumulative histogram and other statistical features. Let's try to use this property for detection of irregularities in textured images.

At first, let's calculate two distributed histograms as two sets of $N(M)$ ordinary histograms (for every column and row of the image pixel matrix):

$$V_i(c) = \{V_{ij}(c)\}, \quad j = 0.255, \ i = 1, N, \tag{1}$$

$$V_j(r) = \{V_{ji}(r)\}, \quad i = 0.255, \ j = 1, M. \tag{2}$$

The distributed histogram shows frequency of pixels intensity values in columns $V_i(c)$ and in rows $V_j(r)$. In the image histograms, the *OX* axis corresponds to the grey level intensities (0–255) in *N* columns (*M* rows) and the *OY* axis corresponds to the frequency of these intensities.

Then let's calculate two distributed cumulative histograms as sets of frequency sums:

$$V_j(cc) = \left\{ \sum_{l=0}^{i} V_{li}(c), \ i = 0.255 \right\}, \quad j = 1, N, \tag{3}$$

$$V_j((r)) = \left\{ \sum_{l=0}^{i} V_{li}(r), \ i = 0.255 \right\}, \quad j = 1, M, \tag{4}$$

where $V_i(cr)$ and $V_j(cc)$ are cumulative histograms in rows and columns, $V_{ij}(c)$ and $V_{li}(r)$ are intensity frequency in column and row, N and M are numbers of columns and rows.

In Fig. 3, *a* schematic example of the distributed histogram is given. Colors of the top planes are proportional to heights of parallelepipeds.
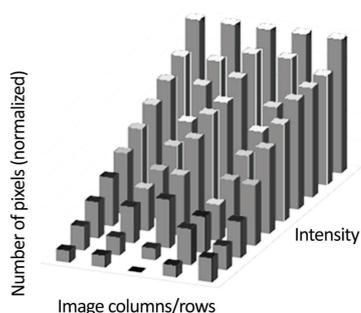
Fig. 3. Distributed cumulative histogram of abstract image

For further processing of the DCH let's present it by a flat *2D* image on the plane *OX*, *OI* is a top view on the three-dimensional distributed histogram in Fig. 3 along the *OI* axis. In the new image each value of the pixel intensity corresponds to the pixel frequency in the columns or rows given by the DCH in Fig. 3:

$$I_i(c) = 255 \times V_{i-1}(cc) / N, \ i = 0.255, \ V_{-1}(cc) = 0, \tag{5}$$

$$I_j(r) = 255 \times V_{j-1}(cr) / M, \ j = 0.255, \ V_{-1}(cr) = 0. \tag{6}$$

In Fig. 4 the DCH for the image in Fig. 1, *a* is presented.

Fig. 4. Distributed cumulative histograms (view from *OX* plane)

In Fig. 4 it is difficult to distinguish a small number of pixels responsible for defects. To make them to be more visible let's fill the closed regions of white and black colors by grey color by the flood-fill algorithm [17]. Therefore, let's get an informative part of the distributed cumulative histogram (IDCH). In Fig. 5 (front side) and Fig. 6 (left side), there are two IDCH with grey backgrounds.

The grey color marks an absence of information. All the other colors stay unchanged.

Fig. 5. Informative distributed cumulative histograms (view from *OX* plane)
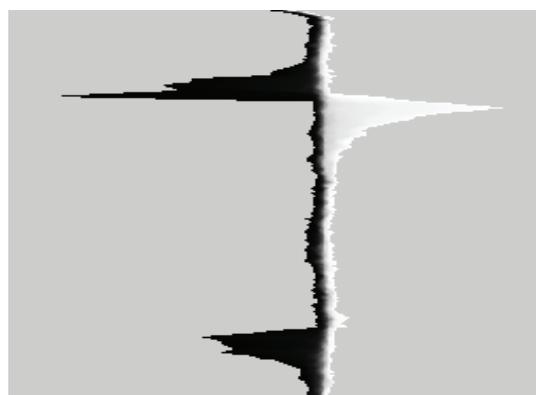
Fig. 6. Distributed cumulative histograms (view from *OY* plane)

3. Research results and discussion

3.1. Measurement of distributed defect intensity by IDCH and statistical features. Defects on the metal surface could differ by size, shape, intensity and other properties. Conditionally let's consider three classes of defects: distributed, concentrated and combined. The metal images from the first class have very small groups of defective pixels scattered on their surfaces. In Fig. 7 an example of creased paper is shown. The second class of defects is characterized by small groups of defective pixels concentrated in some places of the surface. The example of such images in Fig. 1, *a* is given.

For automatic control of the metal samples, it is necessary a procedure in which the defective samples would be neglected automatically. Thus, it is necessary to apply a method for measurement of defect values.

Measurement of dark and light defects directly on the sample in Fig. 7 is not effective because the number of its pixels is very small comparatively to all pixels of the image.
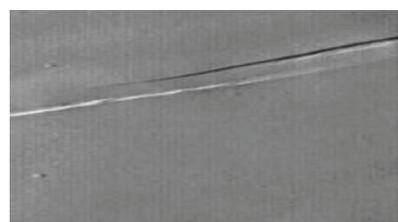
Fig. 7. Sample of the creased paper

That is why the more productive approach is to use the IDCH image (Fig. 8). In this figure it is possible to see many dark and light pixels (excluding background) caused

by small numbers of dark and light pixels from the original sample. In addition, it is possible to see that the left half of the image is more light than right one. Let's neglect this fact since it has a weak influence on the result accuracy.

For a more accurate measurement of white and black defects, let's divide the IDCH into two parts: top and bottom according to the color distribution – white and black (Fig. 9).
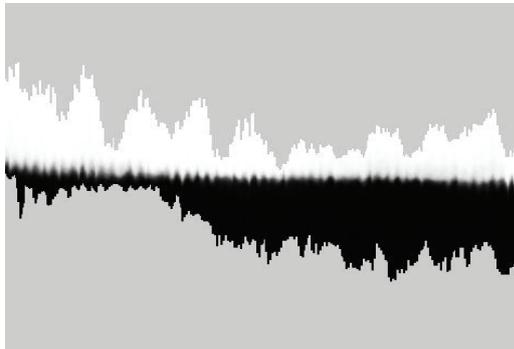
Fig. 8. Informative distributed cumulative histograms of creased paper (view from *OX* plane)

a

b
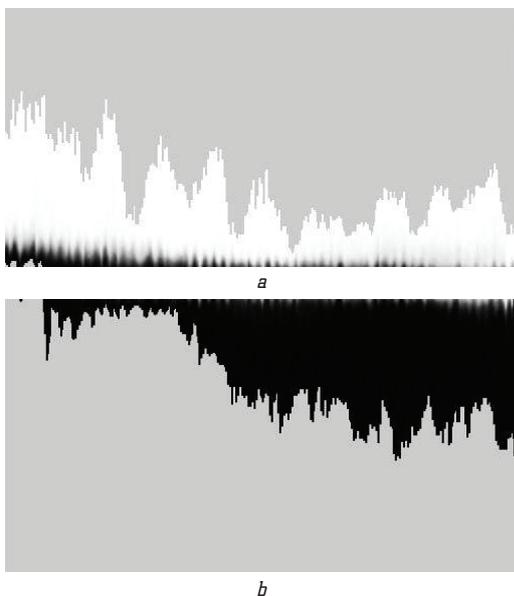
Fig. 9. Two parts of informative distributed cumulative histogram of creased paper: *a* – for white defects; *b* – for black defects

It is necessary to calculate a mean value of pixel intensity in the columns and rows of the image pixel matrix:

$$\bar{I}(j) = \frac{1}{H}\sum_{i=1}^{H} b_{ij}, \qquad (7)$$

$$\bar{I}(i) = \frac{1}{W}\sum_{j=1}^{W} b_{ij}, \qquad (8)$$

where b_{ij} is a pixel intensity in i-row and j-column ($1 \le i \le H$, $1 \le j \le W$), W and H represent the number of columns and the number of rows respectively.

For every part of informative cumulative histogram, let's apply formula (7) for a mean value of pixel intensity in columns of the pixel matrix. As a result, there are two charts in Fig. 10.
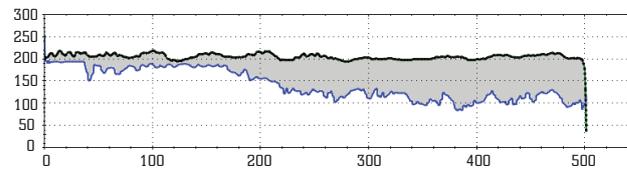
Fig. 10. A mean value of pixel intensity in columns for two parts of informative distributed cumulative histogram

Two charts differ between themselves by marked area. In this area, the distance between the two charts is calculated as the sum of absolute or quadrat values of differences of functions for every column:

$$L_y = \frac{1}{W}\sum_{x=1}^{W} |f_1(x) - f_2(x)|, \qquad (9)$$

$$L_x = \frac{1}{H}\sum_{y=1}^{H} |f_1(y) - f_2(y)|. \qquad (10)$$

Having values of the manufacture tolerance and calculated L_y by the formula (9), the program (or user) takes a decision: to reject or to accept the current sample as suitable.

3.2. Measurement of distributed and concentrated defects intensity by model and statistical features. In Fig. 11 there is an example of metal painting having irregular colors. Measurement of defect intensity directly on this sample is not effective because it is unknown the real correct color. Therefore, it is necessary the etalon metal sample without defects and its informative distributed histogram could be used.

Fig. 11. Metal surface with painting defects

The IDCH of the original sample is shown in Fig. 12. It is possible to see grey background (absence of information), many dark and light pixels caused by pixels of correct and incorrect painting. Colors of the sample are different. Therefore, the border of transition from black to white colors is the curve.
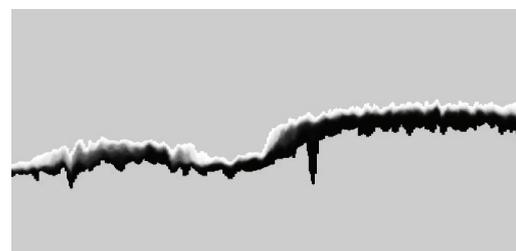
Fig. 12. Informative distributed cumulative histograms of painted metal (view from *OX* plane)

By the image from Fig. 12 it is difficult to say what part of the border is correct. Therefore, it is necessary to

compare it with the IDCH of the correct model of metal sample. The model and its IDCH are shown in Fig. 13.

Let's apply the formula (8) for a mean value of pixel intensity in a row of the image matrix to the following images: the IDCH of original metal sample from Fig. 12 and the IDCH of the model sample from Fig. 13, b. In a result, let's have two charts in Fig. 14.

a

b
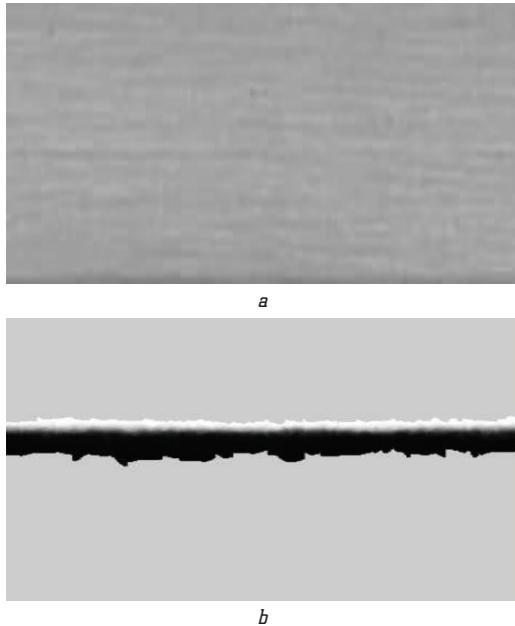
Fig. 13. Images of: *a* – metal sample;
b – informative distributed cumulative histogram
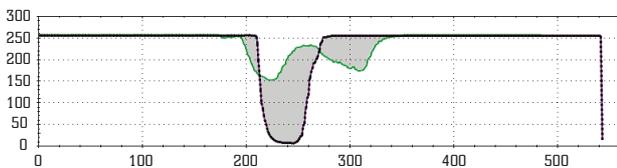
Fig. 14. Mean value of pixel intensity in row: for informative distributed cumulative histogram of the original metal image and model

Two charts differ between themselves by marked area. Let's measure this area. Distances between two charts as the sum of absolute or squared values of difference between the functions for every row by formula (10) are calculated.

Having values of the manufacture tolerance and calculated L_x the program takes a decision: to reject or to accept the current sample as suitable.

3.3. Measurement of concentrated defects coordinates and intensity by clustering. To solve the task, is applied the known hierarchical data clustering approach to rectangles of the image. Function F_{ij} is a weighted sum of modules of differences (Manhattan distance):

$$F_{ij} = \{w_1 \,|\, a_i - a_j \,| + w_2 \,|\, b_i - b_j \,| + w_3 \,|\, c_i - c_j \,| + ...\} / k \cdot r, \quad (11)$$

or a weighted sum of squares (Euclidean distance):

$$F_{ij} = \{w_1 [a_i - a_j]^2 + w_2 [b_i - b_j]^2 + w_3 [c_i - c_j]^2 + ...\} / k \cdot r, \quad (12)$$

where a, b, c are the features that form a cluster in a space; k, r are the numbers of leafs in the i-th and j-th clusters of the tree; w_i is a weight.

The traditional hierarchical data-clustering algorithm consists of the following steps:

S0. For all the points of the input set $x_i, x_j \in X$. At the beginning of the rolling process points x_i, x_j are rectangles (leafs) of the image, than clusters are groups of rectangles.

S1. Search for candidate pairs using the similarity function:

$$\forall (x_i, x_j (j > i)) \quad \text{to count} \quad F(x_i, x_j).$$

S2. Search for pairs with the smallest distance value:

$$F^*(x_i, x_j) = \min F(x_i, x_j), \quad i, j \in I.$$

Then the pair of points x_i, x_j is being united. The new point (cluster) x_{nA} is being created.

S3. Remove points x_i, x_j from the list of candidates.

S4. End (for the all $x_i, x_j \in X$).

The algorithm builds a binary hierarchical rolling tree (dendrogram) of the clusters by the similarity function. To reduce the algorithmic complexity in the step S3 let's combine those pairs of points (clusters) that satisfy the following condition:

$$F(x_i, x_j) \geq F_0 (1 - k_v), \quad (13)$$

where F_0 is the minimum distance at the concrete level of the rolling tree, k_v ($k_v < 1$) is a factor indicating the distance between the candidates to join at the current level of the tree (k_v a speed and the accuracy factor).

The example of the dendrogram illustrating the clustering process is shown in Fig. 15.
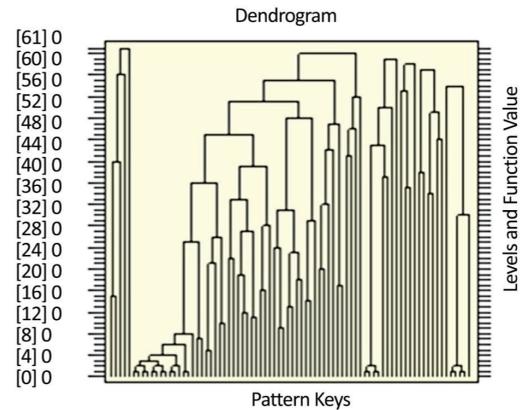
Fig. 15. An example of dendrogram

To illustrate a work of the clustering algorithm let's consider the image of a metal sample with two holes (Fig. 1, *a*). In order to obtain the lowest nodes of the tree (leaves), the input image is divided by the set of horizontal and vertical lines (Fig. 16). For each rectangle the relative value of the full intensity is calculated. The relation is taken to the pixel intensity from full image (all pixels).

After the rolling up process has been performed, one more characteristic for every rectangle – its number of a cluster to which it belongs is obtained. Fig. 17 demonstrates the clustering process and Fig. 18 shows the 6×4 part of the clustered matrix containing one hole. Input data were the metal image, covered by the 10×10 grid and number of clusters was seven. In the image each cluster is marked by a corresponding greyscale color. The clusters with higher intensity are lighter. The image of the metal

sample has dimensions 250×250. So, each rectangle has dimension 25×25.
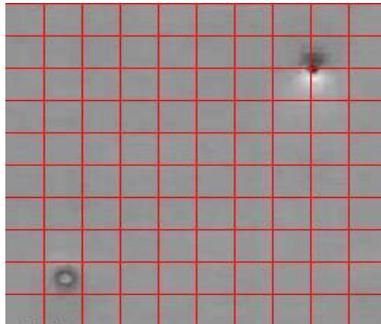
Fig. 16. Coverage of a metal sample by a grid
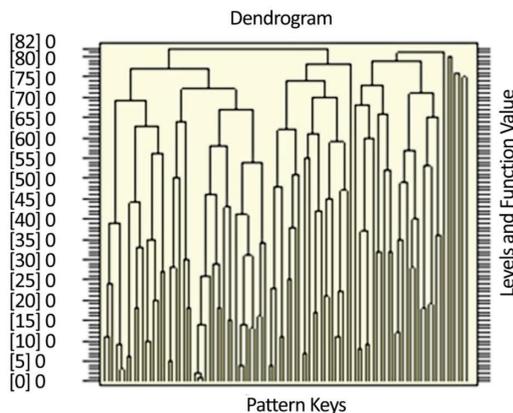
Fig. 17. Clustering dendogram of original image

Elements of seven clusters let's divide into three groups: 6, 7 – responsible for the dark defects; 1, 2, 3 – responsible for the light defects; 4, 5 represent a background.

7	8	9	10
V1.0,009863808 clId.4	V1.0,009856795 clId.4	V1.0,009845681 clId.4	V1.0,01002706 clId.4
V1.0,009850105 clId.4	V1.0,009332087 clId.6	V1.0,008853237 clId.7	V1.0,01004788 clId.4
V1.0,00989294 clId.4	V1.0,01109104 clId.1	V1.0,01122138 clId.1	V1.0,01010777 clId.3
V1.0,009973864 clId.4	V1.0,01043998 clId.2	V1.0,01049501 clId.2	V1.0,01003364 clId.4
V1.0,01003839 clId.4	V1.0,01017272 clId.3	V1.0,01015902 clId.3	V1.0,01008931 clId.3
V1.0,01001961 clId.4	V1.0,01004788 clId.4	V1.0,01004184 clId.4	V1.0,01002501 clId.4

Fig. 18. Part of the clustered matrix of the metal image

In Fig. 3 the value of the element of the first cluster is 0.0111. The value of the element of the seventh cluster is 0.0088. Values of the elements from seven clusters are presented in Table 1. It is possible to see that the difference of values of the elements in the matrix is very small.

Table 1

Values of the elements in clusters for original image

White defects			Background		Black defects	
1	2	3	4	5	6	7
0.0112	0.0100	0.0101	0.0100	0.0096	0.0093	0.0088

In order to increase the distance between the informative defect pixels and the uninformative pixels of the background the image with defects is segmented, that is, the dark part of the background is separated and replaced with the white pixels (Fig. 19). By doing so, let's attempt to increase the range of intensity values of elements in and between clusters covering the image with defects. Applying the clustering algorithm to the transformed image results in the matrix (Fig. 20) which clearly shows the rectangles having defects.
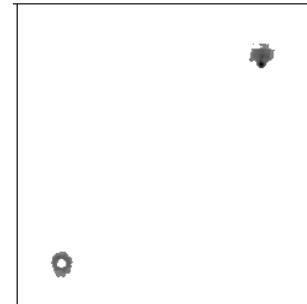
Fig. 19. Segmented image with defects

7	8	9	10
V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1
V1.0,01005 clId.1	V1.0,009341 clId.5	V1.0,008525 clId.6	V1.0,01005 clId.1
V1.0,01005 clId.1	V1.0,009998 clId.3	V1.0,009788 clId.4	V1.0,01005 clId.1
V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1
V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1
V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1	V1.0,01005 clId.1

Fig. 20. Part of the clustered matrix of the segmented image

In Fig. 20 the value of the element of the first cluster is 0.0105 and the value of the element of the seventh cluster is 0.00956. The values of the elements from seven clusters are given in Table 2. The difference of the values of the elements in the matrix is very small too and similar to the matrix of the original image.

Table 2

Values of the elements in clusters for segmented image

Background		Dark defects				
1	2	3	4	5	6	7
0.01005	0.0104	0.0099	0.0097	0.0093	0.0085	0.0080

To perform the next transformation the segmented image is inverted (Fig. 21). The number of white pixels in the segmented image is the same as that of black pixels in the inverted image. The significant difference lies in the fact that the intensity of the black is 0. In this case if pixels from the background are included in the rectangles, they do not affect its integral intensity.

The segmented and inverted image is being clustered with the same 10×10 grid. The clustered matrix (part in Fig. 22) is obtained. The value of the first cluster's element is 0.0321 and the value of the seventh cluster's element is 0.0.
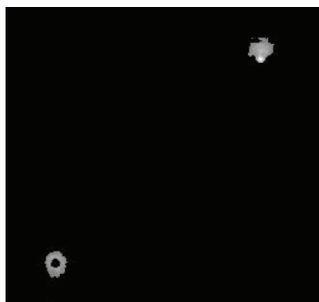
Fig. 21. Segmented and inverted image

Fig. 22. Part of the clustering matrix of the segmented and inverted map

The values of the elements from seven clusters are given in Table 3. The difference of the values of the elements in the matrix is now significant.

The bigger difference of intensity values allows to divide the rectangular area into two and more regions to get more precise intensity features.

Applying a more detailed grid to the surface of the image (for example, 15×15) is possible to see in Fig. 23 how defects from the metal sample are being reflected on clustered matrix. In Fig. 24 the process of clustering is illustrated by the dendogram.

Table 3

Values of the elements in clusters for segmented and inverted image

Dark defects					Background	
1	2	3	4	5	6	7
0.4443	0.3220	0.1505	0.0550	0.0012	0.0080	0.0000

Fig. 23. Clustering matrix of segmented and inverted metal image by the 15×15 grid
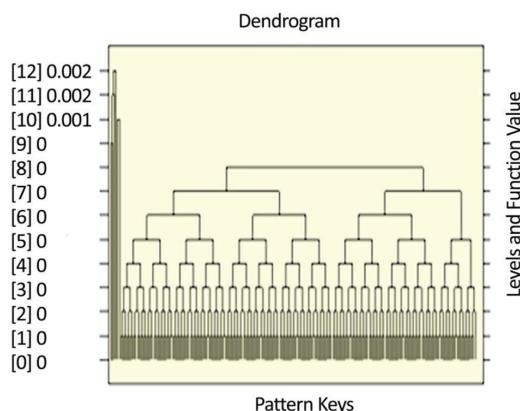
Fig. 24. Clustering dendogram of inverted image

3.4. Measurement of defect intensity by clustering of distributed cumulative histogram. The image of the IDCH by *OX* plane gives a possibility to determine the next features of defects: coordinates and sizes by the axis *OX*, size by axis *OY*, integral intensity as the definite integral of the enveloping function in different intervals etc.

To exclude reciprocal influence of dark and light defects let's divide IDCH into two parts: upper and lower according to color distribution – white and black. The graph of cumulative histogram helps to find the point of division.

Then with two parts of the IDCH there are transformations similar to those earlier performed with the metal image: on the upper part let's change grey color by black and the lower part let's invert and change grey color by black. As it was shown earlier, it is necessary to remote intensity of informative and uninformative pixels.

In result in Fig. 25 let's get two images illustrating sizes and intensity of dark and light defects on the metal surface.

In the previous point the clustering algorithm was used to detect and to rough measurement of defects on the metal surface. Now let's use it to more accurately measure of their intensity. The algorithm is being applied to two parts of the IDCH from Fig. 25. Black color of zero intensity does not affect the intensity value of the rectangles of the image. Therefore, the clustered matrix measures defects by a number of indexed rectangles and intensity value of every rectangle. In sum, this data gives response for the question: to reject or to accept the metal sample.

Let's estimate intensity of defects by indexes K_i of rectangles, which form the closed area of some de-

fect (K_f is the biggest index of cluster, x is the coordinate of the defect):

$$D(x) = \sum_{i=1}^{f} (K_f - K_i), \qquad (14)$$

or by a sum of the rectangle intensity features:

$$S(x) = \sum_{i=1}^{f} I_i. \qquad (15)$$

In Fig. 26 it is possible to see two clustered matrices: for white and black defects. Elements with index 7 are background.
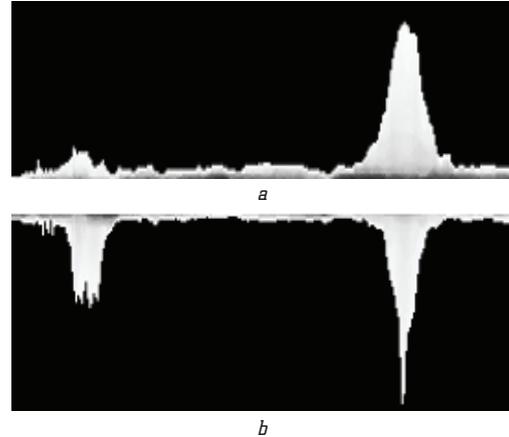
Fig. 25. Two transformed parts of informative distributed cumulative histogram of the metal image: *a* – upper; *b* – lower

V1.0,002468 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,01644 clId.6	V1.0,003509 clId.7	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,02531 clId.5	V1.0,006451 clId.7	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,03251 clId.4	V1.0,01055 clId.7	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,03226 clId.4	V1.0,01688 clId.6	V1.0,006752 clId.7
V1.0,004294 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,03698 clId.4	V1.0,01863 clId.6	V1.0,006752 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,04645 clId.3	V1.0,027 clId.5	V1.0,006752 clId.7
V1.0,005021 clId.7	V1.0,007584 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,05387 clId.2	V1.0,03575 clId.4	V1.0,006752 clId.7
V1.0,04504 clId.3	V1.0,07686 clId.1	V1.0,04382 clId.3	V1.0,04455 clId.3	V1.0,04526 clId.3	V1.0,04634 clId.3	V1.0,04422 clId.3	V1.0,08208 clId.1	V1.0,07593 clId.1	V1.0,0532 clId.2

a

V1.0,05452 clId.2	V1.0,06957 clId.1	V1.0,03564 clId.4	V1.0,02946 clId.4	V1.0,03313 clId.4	V1.0,03119 clId.4	V1.0,03452 clId.4	V1.0,06661 clId.1	V1.0,05513 clId.2	V1.0,04009 clId.3
V1.0,02317 clId.5	V1.0,06331 clId.1	V1.0,0004634 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,05051 clId.2	V1.0,02423 clId.5	V1.0,005964 clId.7
V1.0 clId.7	V1.0,05788 clId.2	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,04161 clId.3	V1.0,01755 clId.5	V1.0,005964 clId.7
V1.0 clId.7	V1.0,04439 clId.3	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,02904 clId.4	V1.0,01095 clId.6	V1.0,005964 clId.7
V1.0 clId.7	V1.0,003841 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,02638 clId.5	V1.0,00662 clId.7	V1.0,005964 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,02098 clId.5	V1.0,001217 clId.7	V1.0,005964 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,02027 clId.5	V1.0 clId.7	V1.0,005964 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,01987 clId.5	V1.0 clId.7	V1.0,005964 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,01591 clId.5	V1.0 clId.7	V1.0,005964 clId.7
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0,01742 clId.5	V1.0 clId.7	V1.0,006816 clId.7

b

Fig. 26. Two clustered matrices of transformed informative distributed cumulative histograms: *a* – upper; *b* – lower

In Fig. 26, *a* the white defect is marked by the next elements: 6, 5, 4, 4, 4, 3, 2 from the eighth column and 6, 6, 5, 4 from the ninth column. Therefore, a value of defect is 19+7=26. In Fig. 26, *b* the black defects are marked by the next elements: 3, 2, 1 from the second column and 5, 5, 5, 5, 5, 5, 4, 3, 2, 5, 5, 5 from the eighth and ninth columns. Therefore, a value of two black defects is 15+30=45.

Calculating intensity there are in the first case $S(x) = 0.307$ and $S(x) = 0.166 + 0.270 = 0.436$ in the second case.

Defect measurement by intensity values is more accurate for the reason that every rectangle contains only information connected with defects.

3.5. Experiments. Some experiments were held with metal (holes, painting defects), paper (creases), glass [16, 18]. They are given in Table 4 with their IDCH images. There are calculated intensity values of white and black defects for every sample by clustering algorithm. To confirm the accuracy of calculation let's use the formula (7) to get the graphs of mean intensity in the columns of transformed both parts of the IDCH images. It is possible to see that the dark and light small anomalies in the samples appear in histograms and could be compared with model samples.

The developed software allows to analyze the images of different materials and textures.

Table 4

Informative distributed cumulative histograms
of defected material samples and their measured intensity

Images of sample	IDCH
1	2

white defect intensity value = 0.710
black defect intensity value = 0.688
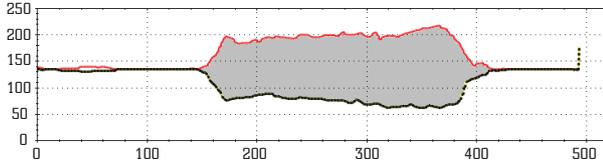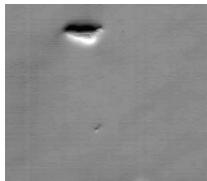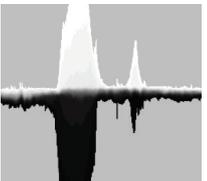
white defect intensity value = 0.571
black defect intensity value = 0.552

1	2

white defect intensity value = 0.212
black defect intensity value = 0.549
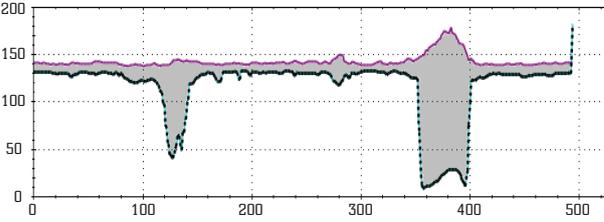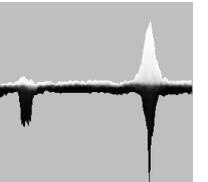
white defect intensity value = 0.307
black defect intensity value = 0.436
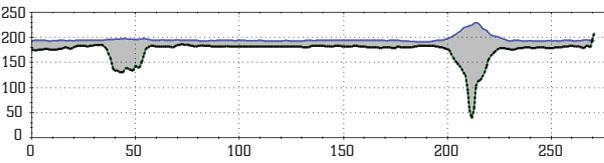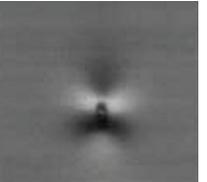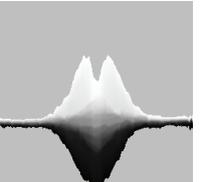
white defect intensity value = 0.482
black defect intensity value = 0.635
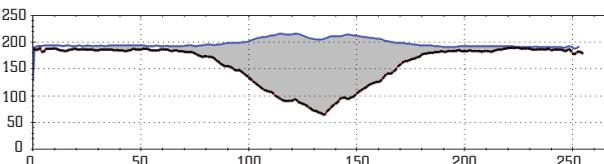
4. Conclusions

The intensity features and coordinates of defects on the sample surface were obtained by clustering algorithm applied to the material image. The informative distributed histogram on a base of distributed histogram is proposed for more precise determination of defects intensity. The IDCH image transformation and clustering algorithm were used for these purposes. The developed software allows to analyzing the images of different materials.

The statistical features of the sample image obtained from the intensity matrix columns and rows have been suggested to calculate the characteristics of informative distributed histograms for estimation of defect intensity.

References

1. Wells, L. J., Shafae, M. S., Camelio, J. A. (2016). Automated Surface Defect Detection Using High-Density Data. *Journal of Manufacturing Science and Engineering, 138 (7)*. doi: http://doi.org/10.1115/1.4032391
2. Ahn, I., Kim, C. (2010). Finding defects in regular-texture images. *16th Korea-Japan Joint Workshop on Frontiers of Computer Vision*. Hiroshima, 478–480.
3. Choi, J., Kim, C. (2012). Unsupervised detection of surface defects: A two-step approach. *IEEE International Conference of Image Processing (ICIP)*. Orlando, 1037–1040. doi: http://doi.org/10.1109/icip.2012.6467040
4. Martins Luiz, A. O., Padua Flavio, L. C., Paulo, E. M. (2010). Almeida automatic detection of surface defects on rolled steel using computer vision and artificial neural networks. *IECON 2010 – 36th Annual Conerence on IEEE Industrial Electronics Society*, 1081–1086. doi: http://doi.org/10.1109/iecon.2010.5675519
5. Jahanbin, S., Bovik, A. C., Pérez, E., Nair, D. (2009). Automatic inspection of textured surfaces by support vector machines. *Optical Inspection and Metrology for Non-Optics Industries*. doi: http://doi.org/10.1117/12.825194
6. Bond, C. (2011). *An efficient and versatile flood fill algorithm for raster scan displays*. Available at: http://www.crbond.com/papers/fldfill_v2.pdf
7. Thilagamani, S., Shanthi, N. (2011). A survey on image segmentation through clustering. *International Journal of Research and Reviews in Information Sciences, 1 (1)*, 14–17.
8. Defects gallery. Available at: http://www.winspection.com/surface-inspection.php
9. Chopade, P. B. (2016). Metal Inspection for Surface defect Detection by Image Thresholding. Available at: https://www.semanticscholar.org/paper/Metal-Inspection-for-Surface-defect-Detection-by-Lohade-Chopade/e321d593df2eab5724f332e6da890d06efd65f25
10. Jeffrey Kuo, C.-F., Peng, K.-C., Wu, H.-C., Wang, C.-C. (2015). Automated inspection of micro-defect recognition system for color filter. *Optics and Lasers in Engineering, 70*, 6–17. doi: http://doi.org/10.1016/j.optlaseng.2015.01.009
11. Ozturk, S. (2017). Detection of PCB Soldering Defects using Template Based Image Processing Method. *International Journal of Intelligent Systems and Applications in Engineering, 4 (5)*, 269–273. doi: http://doi.org/10.18201/ijisae.2017534388
12. Böttger, T., Ulrich, M. (2016). Real-time texture error detection on textured surfaces with compressed sensing. *Pattern Recognition and Image Analysis, 26 (1)*, 88–94. doi: http://doi.org/10.1134/s1054661816010053
13. Singhka, D. K., Neogi, H. N., Mohanta, D. K. (2014). Surface defect classification of steel strip based on machine vision. *International Conference on Computing and Communication Technologies. Hyderabad.* doi: http://doi.org/10.1109/iccct2.2014.7066698
14. Pham, V. H., Lee, B. R. (2014). An image segmentation approach for fruit defect detection using *k*-means clustering and graph-based algorithm. *Vietnam Journal of Computer Science, 2 (1)*, 25–33. doi: http://doi.org/10.1007/s40595-014-0028-3
15. Zheng, K., Chang, Y.-S., Wang, K.-H., Yao, Y. (2016). Thermographic clustering analysis for defect detection in CFRP structures. *Polymer Testing, 49*, 73–81. doi: http://doi.org/10.1016/j.polymertesting.2015.11.009
16. Xu, R., WunschII, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks, 16 (3)*, 645–678. doi: http://doi.org/10.1109/tnn.2005.845141
17. Naz, S., Majeed, H., Irshad, H. (2010). Image segmentation using fuzzy clustering: A survey. *6th International Conference on Emerging Technologies (ICET)*. Islamabad, 18–19. doi: http://doi.org/10.1109/icet.2010.5638492
18. Yi Yang, Dong Xu, Feiping Nie, Shuicheng Yan, Yueting Zhuang. (2010). Image Clustering Using Local Discriminant Models and Global Integration. *IEEE Transactions on Image Processing, 19 (10)*, 2761–2773. doi: http://doi.org/10.1109/tip.2010.2049235

Melnyk Roman, Doctor of Technical Sciences, Professor, Department of Software, Lviv Polytechnic National University, Ukraine, e-mail: ramelnyk@polynet.lviv.ua, ORCID: http://orcid.org/0000-0002-4329-6740

Kvit Roman, PhD, Associate Professor, Department of Mathematics, Lviv Polytechnic National University, Ukraine, e-mail: roman.i.kvit@lpnu.ua, ORCID: http://orcid.org/0000-0002-2232-8678