IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Analysis of Metal Defects by Clustering the Sample and Distributed Cumulative Histogram

Roman Melnyk
Software Department
Lviv Polytechnic National University
Lviv, Ukraine
ramelnyk@polynet.lviv.ua

Yurii Kalychak
Software Department
Lviv Polytechnic National University
Lviv, Ukraine
yurii.i.kalychak@lpnu.ua

***Abstract*—In this paper the clustering algorithm was used to classify the regions of the metal sample with defects to determine their coordinates. The informative distributed cumulative histogram is proposed. To measure sizes and intensity of defects the IDCH image is transformed and clustered.**

Keywords—image intensity, surface, defects, clustering, pixel, segmentation, inversion, distributed cumulative histogram.

I. Introduction

A big number of defect determination method differ between themselves by features and extraction algorithms. The paper [1] considers the probability of detecting size and magnitude of defects in addition to the probability of error 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 amplitude spectrum, and the defect map is then obtained by transforming the difference to spatial domain.

The approach for defect detection in [3] consists of two phases: global estimation and local refinement. First, by applying a spectral-based approach in a global manner roughly estimates defects. Second 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.

Many papers present the image segmentation techniques using clustering [8-12]. For example, the algorithm in [8] uses k-means algorithm to split the original image into regions based on Euclidean color distance to produce an over-segmentation result. In [9] cluster analysis (TCA) method for automatic defect detection is based on three-dimensional image segmentation. Fuzzy, C-Means, K-Means clustering methods [10-13] are the most wide-spread approaches for image segmentation, pattern recognition, finding the optimal segmentation threshold and classification.

The majority of the above-mentioned approaches are quite complicated and time-consuming. In this paper, the clustering algorithm for the calculation of image intensity distribution is developed.

II. Determination of Defect Coordinates by Intensity Clustering

To illustrate a work of the clustering algorithm we consider the image of a metal sample with two holes (Fig. 1a) [5]. 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. 1b). 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 intensity).

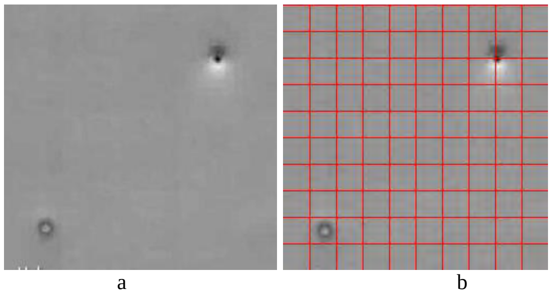


Fig. 1. Metal sample and its coverage by a grid

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. 2 demonstrates the clustering process and Fig. 3 shows the 6x4 part of clustered matrix containing one hole. Input data were the metal image, covered by the 10x10 grid and a number of clusters as seven. In the image each cluster is marked by a corresponding grayscale color. The clusters with higher intensity are lighter. The image of the metal sample has dimensions 250x250. Thus, each rectangle has dimension 25x25.

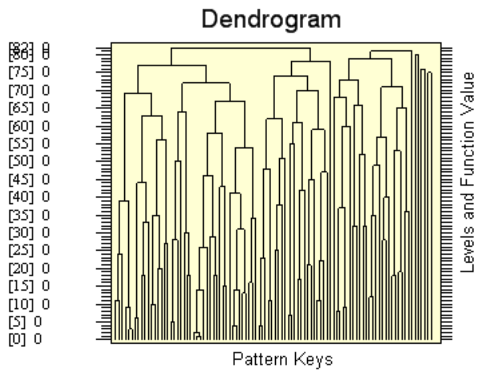
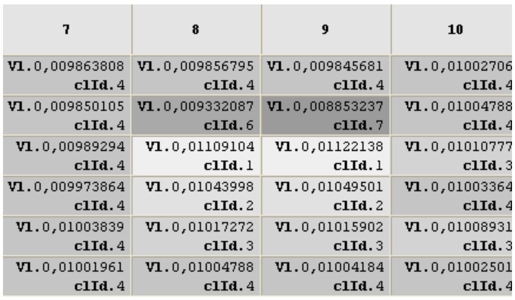


Fig. 2. Dendogram of clustering of original image

Elements of 7 clusters we 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

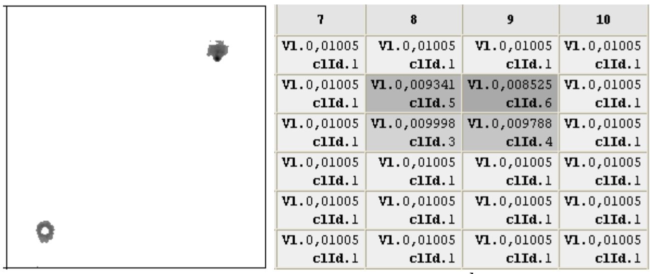
Fig. 3. Part of the clustering matrix of the metal image

In Fig. 3 the value of the element of the first cluster is 0,0111. The value of the element from the seventh cluster is 0,0088. Values of elements from seven clusters are presented in the Table I. We can see that the difference of values of the elements in the matrix is very small.

TABLE I. VALUES OF ELEMENTS IN CLUSTERS (ORIGINAL IMAGE)

White defects			Background		Black defects	
1	2	3	4	5	6	7
0,0112	0,0100	0,0101	0,010	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. 4a). By doing so, we 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. 4a) which clearly shows the rectangles having 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

a b

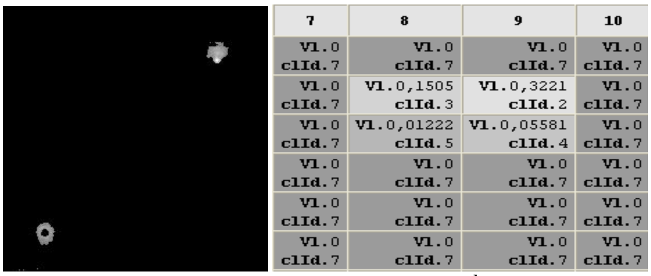
Fig. 4. Segmented image with defects (a) and part of its clustering matrix (b)

In Fig. 4a, the value of the element of the first cluster is 0,0105, and the element value of the seventh cluster is 0,00956. The values of elements from seven clusters are given in the Table II. The difference of values of the elements in the matrix is very small too and similar to the matrix of the original image.

TABLE II. VALUES OF ELEMENTS IN CLUSTERS (SEGMENTED IMAGE)

Background		Dark defects				
1	2	3	4	5	6	7
0,01005	0,0104	0,0099	0,0097	0,0093	0,0085	0,008

To perform the next transformation the segmented image is inverted (Fig. 5a). 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, but the intensity of the white is 255. In this case if pixels from the background are included in the rectangles they do not affect its integral intensity.



7	8	9	10
V1.0 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7
V1.0 clId.7	V1.0,1505 clId.3	V1.0,3221 clId.2	V1.0 clId.7
V1.0 clId.7	V1.0,01222 clId.5	V1.0,05581 clId.4	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 clId.7	V1.0 clId.7	V1.0 clId.7	V1.0 clId.7

a b

Fig. 5. Segmented and inverted image (a) and part of its clustering matrix (b)

The segmented and inverted image is being clustered with the same 10x10 grid. The clusters matrix (part in Fig. 5b) is obtained. The value of the first cluster's element is 0,0321 and the value of the seventh cluster's element is 0,000. The values of elements from seven clusters are given in the Table III. The difference of values of the elements in the matrix is now significant.

TABLE III. VALUES OF ELEMENTS IN CLUSTERS (SEGMENTED AND INVERTED IMAGE)

Dark defects					Background	
1	2	3	4	5	6	7
0,4443	0,322	0,1505	0,055	0,0012	0,008	0,0

A bigger difference of intensity values allows to divide the rectangle areas into two ones: light and dark to get more precise intensity features.

Applying a more detailed grid to the surface of the image (for example, 15x15) we see in Fig. 6 how defects from the metal sample are being reflected on clustered matrix.

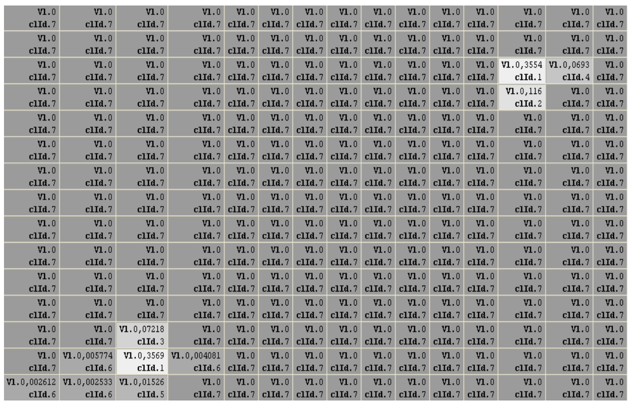


Fig. 6. Clustering matrix of segmented – inverted metal image by the 15x15 grid

III. MESUREMENT OF DEFECT INTENSITY BY CLUSTERING OF DISTRIBUTED CUMULATIVE HISTOGRAM

The previous approach does not allow to measure the defect dimensions and intensity precisely. The reason is that the rectangles of the original image contain dark and light defects and the rectangles of the segmented image does not contain dark or light pixel. So, to get more accurate data reflecting the defects we propose a new approach based on the concept of cumulative histogram namely a distributed cumulative histogram (DCH). We distinguish two types of DCH: view from OX and view from OY.

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

$$V_i(c) = \{V_{ij}(c)\},\ j = 0,255,\ i = 1,N \quad (1)$$

$$V_j(r) = \{V_{ji}(r)\},\ i = 0,255,\ j = 1,M \quad (2)$$

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 shows the gray level intensities in N columns (M rows) and the OY axis shows the frequency of these intensities.

Then we calculate two distributed cumulative histograms as sets of frequency sums:

$$V_j(cc) = \{\sum_{l=0}^{i} V_{li}(c), i = 0,255\},\ j = 1,N \quad (3)$$

$$V_j(cr) = \{\sum_{l=0}^{i} V_{li}(r), i = 0,255\},\ j = 1,M \quad (4)$$

where $V_i(c)$, $V_j(cc)$ – histogram and cumulative histogram in columns, $V_{ij}(c)$ – an intensity frequency in column, N, M – numbers of columns and rows.

A schematic example of distributed histogram is given in Fig. 7.

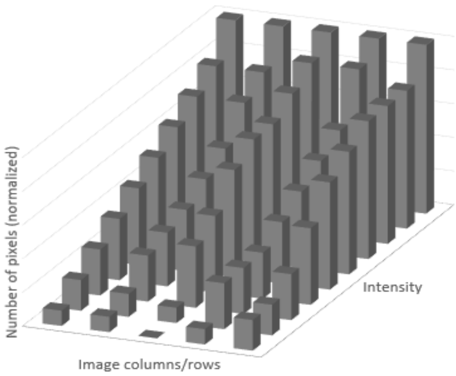


Fig. 7. Distributed cumulative histogram of abstract image

For further processing of DCH we present it by a flat 2D image on the plane OX, OI – a top view on the three-dimensional distributed histogram in Fig. 7 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 DCH in Fig. 7:

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

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

For the image in Fig. 2a the DCH is presented in Fig. 8.

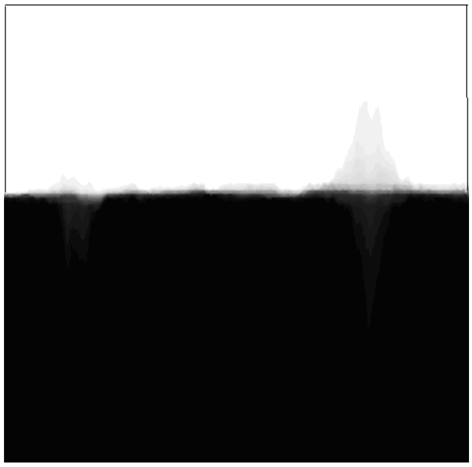


Fig. 8. Distributed cumulative histograms (view from OX plane)

In Fig. 8 it is difficult to distinguish a small number of pixels responsible for defects. To make them to be more visible we fill the closed regions of white and black colors by grey color using the flood-fill algorithm. So, we get informative part of the distributed cumulative histogram (IDCH) in Fig. 9. On it the grey color marks an absence of information. All the other colors stay unchanged.

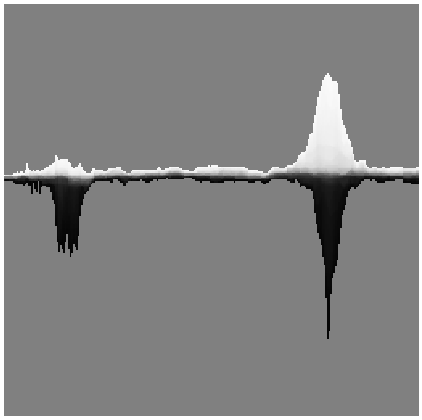


Fig. 9. Informative distributed cumulative histograms (view from OX plane)

The image of the IDCH by OX plane gives us possibility to determine the next features of defects: coordinates and sizes by 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 we divide IDCH into two parts: upper and lower according to color distribution – white and black. The graph of cumulative histogram helps us to find the point of division.

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

In result in Fig. 10 we get two images illustrating sizes and intensity of dark and light defects on the metal surface.

In the previous chapter the clustering algorithm was used to detect and to rough measurement of defects on the metal surface. Now we apply it for more precise measurement of their intensity.

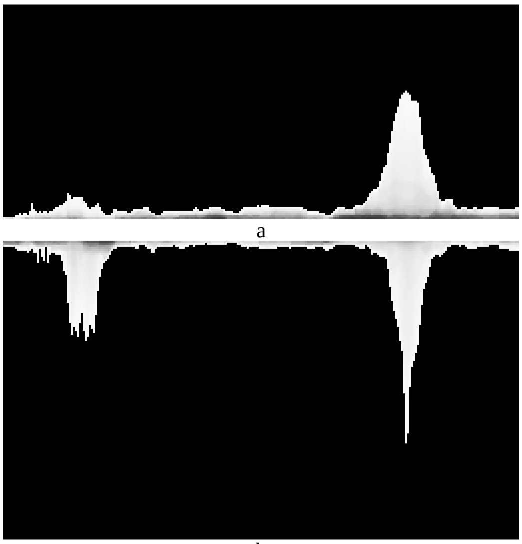


Fig. 10. Two transformed parts of informative distributed cumulative histogram: a – upper, b – lower

The algorithm is being applied to two parts of the IDCH from Fig. 10. Black color of zero intensity does not affect the intensity value of the rectangles of the image. Thus, on the clustered matrix we measure defects by a number of indexed rectangles and intensity value of every rectangles. In sum this data gives us response for the question: to reject or to accept the metal sample.

We estimate intensity of defects by indexes K_i of rectangles which form the closed area of some defect:

$$D(x) = \sum_{i=1}^{f} (K_f - K_i) \qquad (7)$$

where K_f is the biggest index of cluster, x – coordinate of defect. Or by a sum of the rectangle intensity features:

$$S(x) = \sum_{i=1}^{f} I_i \qquad (8)$$

Let us calculate a mean value of a pixel intensity in the j-th columns of the image pixels matrix:

$$I(j) = 1/H \sum_{i=1}^{H} b_{ij} \qquad (9)$$

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

IV. EXPERIMENTS

In Fig. 11 we see two clustered matrices: for white and black defects. Elements with index 7 represent background. In Fig. 11a the white defect is marked by the elements with next indices: 6, 5, 4, 4, 4, 3, 2 from the eighth column and 6, 6, 5, 4 from the ninth column. So, a value of defect is 19+7=26. In Fig. 11b the black defects are marked by the elements with next indices: 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. So, a value of two black defects is 15+30 =46.

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

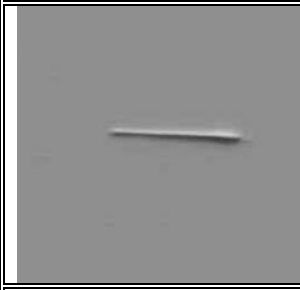
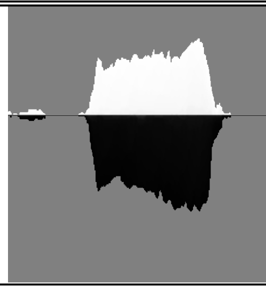
b

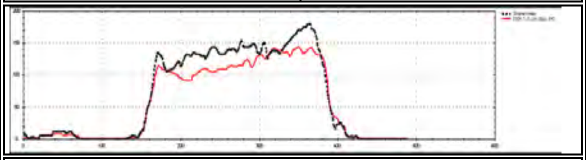
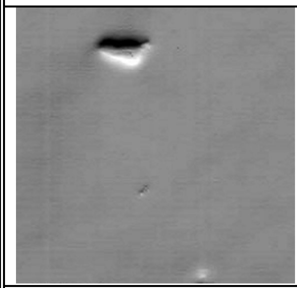
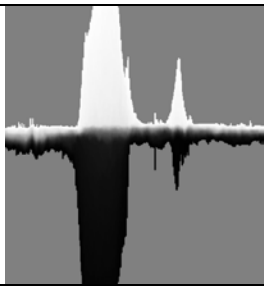
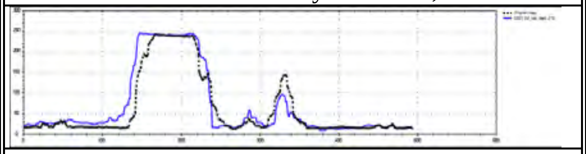
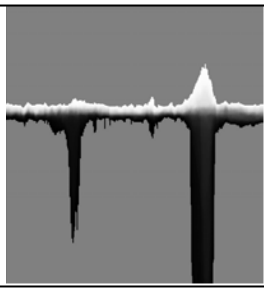
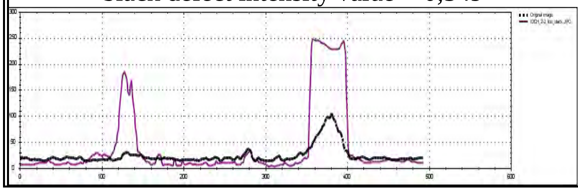
Fig. 11. Two clustered matrixes of transformed informative distributed cumulative histograms: a – upper, b – lower

Calculating intensity we have in the first case $S(x) = 0{,}307$ and in the second case $S(x) = 0{,}166 + 0{,}270 = 0{,}436$. Defect measurement by intensity values is more accurate for the reason that every rectangle contains only information connected with defects.

Some experiments were held with metal (scratches and holes), paper (creases) [6, 7]. They are given in Table IV with their IDCH images. Also in the Table IV calculated intensity values of white and black defects for every metal sample are given. To confirm the accuracy of calculation we use the formula (9) to get the graphs of mean intensity in the columns of transformed IDCH images. We can see that dark and light small anomalies in the metal samples appear in histograms and could be compared with etalon samples.

TABLE IV. INFORMATIVE DISTRIBUTED CUMULATIVE HISTOGRAMS OF DEFECTED MATERIAL SAMPLES

<table>
<tr><th>Images of metals</th><th>IDCH</th></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">white defect intensity value = 0,710
black defect intensity value = 0,688</td></tr>
</table>

Images of metals	IDCH
	
	
white defect intensity value =0,571 Black defect intensity value = 0,552	
	
	
white defect intensity value =0,212 black defect intensity value = 0,549	
	

V. CONCLUSION

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

REFERENCES

[1] L.J. Wells, M. S. Shafae and J.A. Camelio, "Automated Surface Defect Detection Using High-Density Data," *J. Manuf. Sci. Eng.*, 138(7), Mar, 2016.

[2] I. Ahn and Ch. Kim, "Finding Defects in Regular-Texture Images," *16th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Hiroshima, Japan, pp. 478-480, Feb. 2010.

[3] J. Choi and Ch. Kim, "Unsupervised Detection of Surface Defects: A Two-Step Approach," *IEEE International Conference of Image Processing (ICIP)*, Orlando, USA, pp. 1037-1040, Sep. 2012.

[4] L.A.O. Martins, F.L.C. Padua, and P.E.M. Almeida, "Automatic Detection of Surface Defects on Rolled Steel Using Computer Vision and Artificial Neural Networks IECON," 36th *Annual Conference on IEEE Industrial Electronics Society*, pp. 1081-1086, 2010.

[5] S. Jahanbina, A.C. Bovika, E. Perezb, and D. Nair, "Automatic Inspection of Textured Surfaces by Support Vector Machines," [Electronic resource] Link: ttps://live.ece.utexas.edu/publications

[6] Wintriss defects gallery [Electronic resource] Link: http://www.weco.com/surface-inspection

[7] Defect detection of various films [Electronic resource] Link: http://cilabs.kaist.ac.kr/research/image-analysis/defect-detection

[8] V.H. Pham, and B.R. Lee, "An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm," Vietnam Journal of Computer Science, vol. 2, iss. 1, pp 25–33, February 2015,

[9] K. Zheng, Y.-S. Chang, K.-H. Wang, and Y. Yao, "Thermographic clustering analysis for defect detection in CFRP structures," Polymer Testing, vol. 49, pp. 73-81, February 2016.

[10] R. Xu, and D. Wunsch, "Survey of clustering algorithms", IEEE Transactions on Neural Networks, vol. 16, iss. 3, pp. 645 – 678, May 2005.

[11] S. Naz, H. Majeed, and H. Irshad, "Image segmentation using fuzzy clustering: A survey," 6th International Conference on Emerging Technologies (ICET), pp. 181 – 186, 18-19 Oct. 2010.

[12] S. Thilagamani1 and N. Shanthi, "A Survey on Image Segmentation Through Clustering," International Journal of Research and Reviews in Information Sciences, vol. 1, no. 1, pp. 14-17, March 2011.

[13] Y. Yang, D. Xu, F.Nie, S. Yan, and Y. Zhuang, "Clustering Using Local Discriminant Models and Global Integration," IEEE Transactions on Image Processing, vol. 19, iss. 10, pp. 2761 – 2773, Oct. 2010

.