UDC 303.732.4 + 004.67 + 004.8

AUTOMATIC CLASSIFICATION OF PAINTINGS BY YEAR OF CREATION

Martynenko A. A. – Senior Lecturer of Department of Software Engineering, Dnipro University of Technology, Dnipro, Ukraine.

Tevyashev A. D. – Dr. Sc., Professor, Head of Department of Applied Mathematics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Kulishova N. E. – PhD, Assistant Professor, Professor of Department of Media Systems and Technologies, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Moroz B. I. – Dr. Sc., Professor, Corresponding Member of the Academy of Applied Electronics, Professor of the Software Engineering Department, Dnipro University of Technology, Dnipro, Ukraine.

Sergienko A. S. – Student of Department of Applied Mathematics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. The problem of automatic verification of the legitimacy of the export of works of art is considered.

Objective. A method is proposed for automatically determining the age of a painting from a digital photograph using a classification that is performed by an intelligent decision-making system.

Method. It is proposed to use the attribute of picture year of creation as the main criterion for making a decision during the customs check of exports legitimacy. Instead of a long and expensive museum examination, photographing works of art in customs conditions and processing photos using a set of descriptors is used. The set of descriptors is proposed, include local binary patterns, their color modification, Haralik's texture features, the first four moments, Tamura's texturt features, SIFT descriptor. The data obtained as a result of descriptors action give the values of several dozen private attributes. They form data vectors, which are then concatenated into a generalized object description vector. In the feature space thus created, automatic classification by weighted k-nearest neighbors is performed. The proposed algorithm calculates the distance between objects in a multidimensional space of attribute values and assigns new objects to already formed classes. The criterion for creating classes is the age of the painting from the existing database. As a measure of the objects proximity, it is proposed to use the Euclid and Minkowski metrics. The calculation of weights for the proposed classification algorithm is performed by the Fisher method.

Results. The effectiveness of the proposed method was investigated in the course of experiments with an image database containing photos of paintings by world, European and Ukrainian artists. Algorithm configuration parameters that provide high classification accuracy are found.

Conclusions. The performed experiments have shown the effectiveness of the selected descriptors for the formation of vector descriptions of images of paintings. The greatest accuracy is provided by descriptor merging, which reveals significant differences in the structural properties of images. This approach to the description of objects in combination with the proposed classification algorithm and the chosen main criterion ensures high accuracy of the obtained solutions. The direction of further research may include the use of convolutional neural networks to improve the accuracy of classification under the condition of a static database.

KEYWORDS: intelligent decision-making system, automatic classification, k-nearest neighbors, image descriptors, feature vector, customs examination, paintings.

ABBREVIATIONS

SOM are self-organizing maps;

SVM are support vector machines;

k-NN is a *k*-nearest neighbors method;

CNN is a convolutional neural network;

LBP are local binary patterns;

SIFT is a scale-invariant feature transform descriptor;

RGB LBP are local binary patterns in RGB color space;

Color LBP are local binary patterns found in color channels.

NOMENCLATURE

g is a pixel brightness;

$s(\bullet)$ is a the Heaviside step function;

P is a size of pixel neighborhood;

g_c is a value of brightness of neighborhood central pixel;

g_p is a brightness value of *p*-th pixel of neighborhood with the size *P*;

μ is a pixel brightness average value;

σ^2 is a pixel brightness deviation from the average;

μ_n is a *n*-th order central moment of random pixel brightness distribution;

μ_3 is a 3-th order central moment or asymmetry of random pixel brightness distribution;

μ_4 is a 4-th order central moment of random pixel brightness distribution;

I is an image;

$P(i,j)$ is a contingency matrix;

i,j are pixel coordinates;

C_H is an image contrast;

DOI 10.15588/1607-3274-2022-2-8

$Corr_H$ is an image correlation;

$Entropy_H$ is an image entropy;

$Energy_H$ is an image energy;

M_B is an image palette redundancy;

$H_{\max}$ is a maximum image entropy;

H_{RGB} is an entropy, calculated for individual *R*, *G*, *B* channels;

A_k is an average value of pixel brightness in neighborhood;

E_k is a texture roughness;

C_k is a texture contrast;

α_4 is a kurtosis;

$H_{dir}(a)$ is a quantized edge directions;

D_k is a histogram of quantized edge directions;

n_{peaks} is a histogram peaks number;

a_p is a peak angular direction;

r is a coefficient that depends on quantization levels of angles;

L_k is a linear similarity;

R_k is a texture regularity;

$\sigma_{coarseness}$ is a standard deviation of texture coarseness;

$\sigma_{contrast}$ is a standard deviation of texture contrast;

$\sigma_{directionality}$ is a standard deviation of texture directionality;

$\sigma_{linelikeness}$ is a standard deviation of texture linelikeliness;

$d(x_i, x_j)$ is a measure of similarity between objects, equal to metric distance between data points;

x_i, x_j are objects to be compared;

f_i is an attribute of object matching;

$\mathbf{c}_i$ is an attributes value;

d_E is an Euclidean metric;

d_M is a Minkowski metric.

INTRODUCTION

Painting has long ceased to be art for the elite – reproductions of paintings can be found on items of clothing, bags, in the form of curtains, as graffiti on the walls of buildings. Such popularization undoubtedly leads to the fact that the originals of paintings are constantly growing in value and have long since turned from objects of art into an accumulating value means. This raises many problems for customs services – export of valuable paintings undermines the economic security of the state.

Verifying the authenticity and value of art objects when crossing state borders is an important, urgent and difficult task. The procedure for exporting cultural property during customs control for examination, organization of expertise and other aspects are regulated by the 1970 UNESCO Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, 1995 UNIDROIT Convention on Stolen or Illegally Exported Cultural Objects, 1954 Europe Cultural Convention [1–3]. In particular, according to the approved procedure, the export of cultural property is possible only if it is confirmed by certificate for the right to export, issued by the Department for the movement of cultural property of the Department of Museum Affairs and Cultural Property under the Ministry of Culture of Ukraine**.** The paintings authenticity takes place during expertise carried out by qualified historians and art critics for a fairly long time. However, it is not uncommon for malefactors to deliberately hide true value of paintings for export, passing them off as much less valuable and therefore do not require a certificate for the right to export. Then the customs service is faced with the need to quickly and accurately assess whether the picture being transported can be classified as a cultural property or not. According to regulatory documents, antiques are items over 100 years old. That is, an operational customs check when exporting paintings abroad is reduced to determining the painting age. The most reliable techniques for this use X-ray fluorescence analysis, infrared and ultraviolet spectroscopy and other methods of analysis. Unfortunately, all of them are now absent in customs arsenal, as well as specialists of corresponding qualifications. At the same time, organizing photographing a picture using a digital camera is a solution that is affordable both in cost and in terms of technical capabilities. An intelligent decision-making system [4, 5] provides painting automatic identification by painting photo and establishing its authenticity and value. Obviously, for operational customs control during the paintings export, it is enough to estimate the painting age and, based on this information, make a decision on export possibility or impossibility.

The object of study is a decision-making process for permission to export paintings during a customs check, which is implemented in an intelligent decision-making system.

The subject of study are methods for automatic classification of paintings images based on a generalized description of their properties with the year of creation as a key attribute.

The main purpose of the work is automatically determining the age of a painting from a digital photograph during classification performed by an intelligent decision-making system.

1 PROBLEM STATEMENT

Suppose given a set of images $X = \{x_1, x_2, ..., x_N\}$, N – number of them. Every image could be described by several characteristics f_i, $i = \overline{1,m}$, their values $\mathbf{c}_i = \left(c_{f_i 1}, ..., c_{f_i k}, ..., c_{f_i n}\right)$, $k = \overline{1,n}$ are results of some

DOI 10.15588/1607-3274-2022-2-8

image processing descriptors. One image characteristic – a key attribute $Y_l, l = 1,...N$ – is known an advance. According to attribute Y set X is marked by some class labels $Z = \{z_1, z_2, ..., z_M\}$. The same label assigns to images x_i, x_j, that have a distinction $d(x_i, x_j) \le \varepsilon$ where ε – similarity level.

The mathematical problem is to find a classification rule $g(X,Y): X \to Z$, that for a given image x_{N+1} minimizes a distinction measure $d\left(x_{N+1}, x_j \mid \forall x_j : g(x_{N+1}, y_{N+1}) = g(x_j, y_j),\ j = 1,...N\right)$.

2 REVIEW OF THE LITERATURE

With development of computing power, which made it possible to process digital photos in high resolution to analyze of high-dimensional data, scientists and engineers began to solve the problem of automatically classifying works of art by photo. These studies use a machine learning approach and are ongoing [6–10]. For pictures automatic classification the most widely used methods of self-organizing maps (SOM) [11], support vector machines (SVM) [12], *k*-nearest neighbors (*k*-NN) [13–16]. Their undoubted advantages are high classification accuracy, ability to fast updating of training datasets, a high learning rate.

Convolutional neural networks (CNN), which have proven their high efficiency in a wide range of tasks related to processing of images of various kinds, are in serious competition for mentioned techniques. CNNs provide higher accuracy compared to other machine learning methods and are fast. Many works demonstrate that application of CNN to automatic classification of picture photos gives positive results [17–21]. However, such networks have a number of disadvantages that limit their use for expeditious customs inspection of paintings for their export possibility. These disadvantages include need for hundreds of thousands or millions of objects for convolutional networks training; training duration, which will increase significantly if it is necessary to update the set of training samples. These factors make convolutional networks computationally and financially costly.

In overwhelming majority of works, automatic classification of paintings is carried out according to several main criteria: by the artist name; by the artistic style or genre to which work can be attributed. This is necessary when identifying and confirming paintings authenticity. Prompt check of painting items value at customs lead to the need to classify paintings by age. There are not so many such works, since dozens of genres can be represented in painting at the same time. Nevertheless, researchers do not abandon this attribute, successfully including it in paintings automatic categorization systems [17].

In this paper, it is proposed to use weighted *k*-nearest neighbors algorithm, which has successfully proven itself in solving complex problems of image classification and allows to successfully updating the dataset for training, and use the picture age as the main classification criterion.

3 MATERIALS AND METHODS

The classification accuracy depends on choice of attributes characterizing objects. When working with images, such algorithms and descriptors as Local binary patterns (LBP) [22] and their color modifications; Haralik's texture features [23]; SIFT descriptor [24] have successfully proven themselves.

Local binary patterns (LBP) [22] – descriptors describing properties of neighborhoods of a given pixel in the image:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \tag{1}$$

where $s(\bullet)$ – the Heaviside step function *step* (*x*), which returns 0 if $x \ge 0$, and 1 otherwise.

Based on results of comparing brightness g_c and g_p, a histogram for each pixel is built. These histograms are normalized, compressed and combined into a single data vector. The method turned out to be extremely effective, especially in the tasks of separating object from background. One of its modifications – RGB-LBP – allows color images processing. In this case, local binary pattern is calculated in RGB space for each color component separately, and then descriptions are combined. In general, LBPs can be defined in any color space, such modifications are known as Color LBPs.

In analysis practice, an image is often considered as a random process characterized by a certain law of distribution of pixel brightness g as a random variable. The main parameters describe this random variable are the mathematical expectation, variance, and central moments of brightness distribution.

The mathematical expectation in the case of images with a finite number of pixels P is represented by its approximation – the average value:

$$\mu = \lim_{p \to \infty} \frac{1}{P} \sum_{p=1}^{P} g_p . \tag{2}$$

Dispersion allows estimating degree of pixel brightness possible values deviation from the average:

$$\sigma^2 = \sum_{p=1}^{P} (g_p - \mu)^2 . \tag{3}$$

The central moment of the *n*-th order of random variable distribution in the general case for an image can be estimated using the relation:

DOI 10.15588/1607-3274-2022-2-8

$$\mu_n = \sum_{p=1}^{P} \left(g_p - \mu\right)^n . \qquad (4)$$

The first central moment is equal to zero, the second is equal to the variance. The third central moment μ_3 is asymmetry. It demonstrates the asymmetry of the probability density function about the mean. The fourth central moment μ_4 characterizes the sharpness of the probability density function top. Thus, combination of these four indicators – pixel brightness average value, variance, third and fourth central moments – exhaustively describes properties of distribution function of pixel brightness for a specific image. Therefore, they are called "first four moments" and are often used in image analysis problems solving.

Haralik's texture features [23] also describe brightness values statistical properties and are calculated based on the contingency matrix:

$$P(i,j) = \frac{\neq\left[\left(g_1, g_2\right) \in I \mid \left(g_1 = i\right) \wedge \left(g_2 = j\right)\right]}{\neq I} . \qquad (5)$$

where g_1, g_2 – pixels belonging to the image I. Then contrast is found in accordance with the expression

$$C_H(x,y) = \sum_{i,j} (i-j)^2 P(i,j) ; \qquad (6)$$

the correlation is calculated as

$$Corr_H(x,y) = \sum_{i,j} \frac{\left(i-\mu_i\right)\left(j-\mu_j\right) P(i,j)}{\sigma_i \sigma_j} , \qquad (7)$$

entropy:

$$Entropy_H(x,y) = \sum_i \sum_j P(i,j) \log_2 P(i,j) , \qquad (8)$$

energy:

$$Energy_H(x,y) = \sum_i \sum_j P(i,j)^2 . \qquad (9)$$

The important information about images is clearly related to color data. To generalize them, such an indicator as palette redundancy is used [25]:

$$M_B = \frac{H_{\max} - H_{RGB}}{H_{\max}} . \qquad (10)$$

where $H_{\max}$ is maximum image entropy, which for 8-bit color coding is 8 * 3 = 24; H_{RGB} – entropy, calculated by (8), for individual *R*, *G*, *B* channels.

Each image can be viewed as a texture formed by a collection of some repeating and non-repeating elements. The well-known Tamura features effectively describe the texture properties. They include roughness, contrast, directionality, linearity, roughness, and regularity.

The texture roughness characterizes dimensions of main details that form the image. Its estimate is based on calculation of average values within pixels neighborhood:

$$A_k(x,y) = \sum_P \frac{g(i,j)}{2^{2P}} , \qquad (11)$$

where $g(i,j)$ – brightness of pixel with coordinates i, j; *P* is the size of the neighborhood; the texture roughness is then

$$E_k(x,y) = A_k(x,y) - A_k(x',y), x' \neq x . \qquad (12)$$

The texture contrast is estimated based on the fourth moment μ_4 relative to mathematical expectation and variance σ^2 within the neighborhood:

$$C_k(x,y) = \frac{\sigma}{\left(\alpha_4\right)^{0.25}} . \qquad (13)$$

where $\alpha_4 = \frac{\mu_4}{\sigma^4}$ – kurtosis.

The texture directivity is estimated based on a histogram of quantized edge directions $H_{dir}(a)$:

$$D_k(x,y) = 1 - rn_{peaks} \sum_p \sum_{a \in w_p} \left(a - a_p\right)^2 H_{dir}(a) , \qquad (14)$$

where n_{peaks} – histogram peaks number; a_p – peak angular direction; r – coefficient that depends on quantization levels of angles a_p; $a_p = \arctan\frac{\Delta x}{\Delta y}$ are calculated with Prewitt contour detector.

Linear similarities $L_k(x,y)$ are evaluated as average coincidence of edge directions that coincide in pairs of pixels separated by a distance along edge direction in each pixel.

Texture regularity is a generalized feature defined as

$$R_k(x,y) = 1 - r\left(\sigma_{coarseness} + \sigma_{contrast} + \sigma_{directionality} + \sigma_{linelikeness}\right) , \qquad (15)$$

where $\sigma_{coarseness}, \sigma_{contrast}, \sigma_{directionality}, \sigma_{linelikeness}$ are standard deviations for each feature.

Roughness summarizes the contrast and roughness of texture as follows:

DOI 10.15588/1607-3274-2022-2-8

$$Roughness_k(x,y) = E_k(x,y) + C_k(x,y). \quad (16)$$

The well-known SIFT descriptor [24] collects information about the statistics of local directions of pixel brightness gradient. It is stable to shifts, rotations and scale transformations. In problems of image classification, these properties of descriptor turn out to be indispensable, since they allow comparing objects regardless of differences in size, orientation and location in image.

Vectors obtained as a result of separate descriptors using are combined into one common vector for describing the object.

The simplest metric classification method determines the similarity between data points using the chosen similarity measure, and based on this information, assigns new data points to one or another existing class. The algorithm refers to supervised learning methods, is distinguished by its implementation simplicity and rather high performance if data attributes number is small, and classification objects number does not exceed 10^3. The Euclidean distance is used as a measure of similarity:

$$d_E(x_i,x_j) = \sqrt{\left(c_{f_i 1} - c_{f_j 1}\right)^2 + \ldots + \left(c_{f_i n} - c_{f_j n}\right)^2} = \left\| \mathbf{c}_i - \mathbf{c}_j \right\|, \quad (17)$$

where $d(x_i, x_j)$ – measure of similarity between objects, equal to metric distance between data points, x_i, x_j are the objects to be compared, f_i, $i = \overline{1,m}$ are the attributes of object matching, $\mathbf{c}_i = \left(c_{f_i 1}, \ldots, c_{f_i k}, \ldots, c_{f_i n}\right), k = \overline{1,n}$ are the attributes values.

Minkowski metric

$$d_M(x_i,x_j) = {}^{1/p}\!\sqrt{\left(c_{f_i 1} - c_{f_j 1}\right)^p + \ldots + \left(c_{f_i n} - c_{f_j n}\right)^p} = \left\| \mathbf{c}_i - \mathbf{c}_j \right\|^p \quad (18)$$

also extremely useful in image classification tasks.

In this paper, we consider a system for which the features number is large enough. A weakness of weighted *k*-nearest neighbors method is that when you add up a large number of dissimilarities between data points, the sums can be approximately equal. Because of this, classification objects become poorly distinguishable in selected feature space. To make features more distinguishable, use weights assigned to attributes or data points. The simplest solution is to assign the weight value to reciprocal of distance between the points.

In a multidimensional data space, nearest neighbor search can be performed in different ways also. Known modifications suggest dividing the space by hyperplanes, as in k-d tree algorithm [26]. The modification provides high algorithm performance if number of attributes does not exceed 20.

For problems with a large number of dimensions of data space, so-called BallTree algorithm is used [27]. In this case, space is divided into hyperspheres with centers at data points. The known distance between current data point and the centroid of hypersphere allows defining boundaries of distances to all points within hypersphere. This approach reduces time needed to find the nearest neighbor and is most effective for highly structured data, even with very large space dimensions.

4 EXPERIMENTS

To research the approach effectiveness, a set of images of paintings by 50 famous artists who lived at different times, from 15th century to mid-20th century, was used [28]. The set objects are characterized by such characteristics as artist name, years of life, genres, nationality, biographical facts. For artists who have searched for style in their work, there is information about several genres related to the same period of life. Undoubtedly, each of characteristics of picture description can act as a target attribute in classification. In this work, the attribute of artist's lifetime is chosen as the target feature. The total number of images in dataset was 1169. The number of works for studied artists is shown in Table 1.

5 RESULTS

In the first part of experiment, it was studied the influence of descriptor choice on data classification accuracy. Descriptors LBP (1), Color LBP; the first four moments, calculated by (2)–(4); Haralik parameters calculated by (5)–(10); Tamura texture features, estimated by (11)–(16); SIFT descriptor. Examples of original images processing using selected descriptors are shown in Fig. 1, 2.

Applying descriptors to photos of pictures from a dataset gives feature vectors of different dimensions. For example, color LBP gives a feature vector with dimensions 512×1, SIFT descriptor – a feature vector with dimensions 788×128. To solve the classification problem, all feature vectors must be converted into columns, so the final size of the feature vector for SIFT was 100864×1. A generalized feature vector is formed from such vectors by concatenation. The results of classification using single feature vectors for each descriptor and a generalized vector are presented in Table 2.

Table 1 – The number of paintings images included in studied dataset, depending on artist name

Artist name	Modigliani	Kandinsky	C. Monet	Rivera	Magritte	Dali	Klimt
Number of paintings images by author	193	87	73	70	194	138	117

DOI 10.15588/1607-3274-2022-2-8

6 DISCUSSION

The examples (Fig. 1, 2) show that the proposed descriptors unambiguously and fully reflect the structural, morphological and color paintings properties. LBP, color LBP and the first 4 moments reveal the contours and fine details of images, the Haralik features segment the images according to textural characteristics with high accuracy, and the SIFT descriptor finds the feature points in contours.

During the experiments, the hypothesis that one descriptor is enough to accurately classify the picture according to the creation time based on received information was tested.

This hypothesis is explained by the need to provide high algorithm performance simultaneously with high accuracy. The longer image attributes vector, the lower the classifier speed.

The data in Table 2 shows that the use of only one descriptor provides a low quality classifier: solution accuracy varies from 62% to almost 73%. Combining the data vectors received from several descriptors into one vector made it possible to increase the accuracy of the solution by almost 10%, bringing it to 82.71%. Since during the customs check it is not expected that new images will arrive in the stream at a speed comparable to video, we can conclude that the use of generalized description vectors turned out to be a very effective solution.

In the second part of the experiment, it was necessary to find the settings of the classifier that implements the *k*-nearest neighbors method. We checked such settings as a search tree creating algorithm (k-d tree, Balltree), metric (Euclidean, Minkowski), method for calculating weights (inverse to distance, Fisher score), number of neighboring points when deciding whether to belong to a class and size leaves in the search tree.

For the considered problem, attributes properties were such that the Euclidean metric provided sufficient distinguishability of data points. The highest accuracy of the classifier is provided by settings that are given in Table 3.

The artistic manner of each artist influences the classification result. For several of most famous masters in the third part of the experiment, categorization by age was performed. The results are shown in Table 4.

The listed artists worked in the late 19th and early 20th centuries, but their artistic style varies greatly. Despite the dissimilarity of style, the proposed algorithm carried out the classification by the paintings creation time for these masters with high accuracy – 80–82%.

CONCLUSIONS

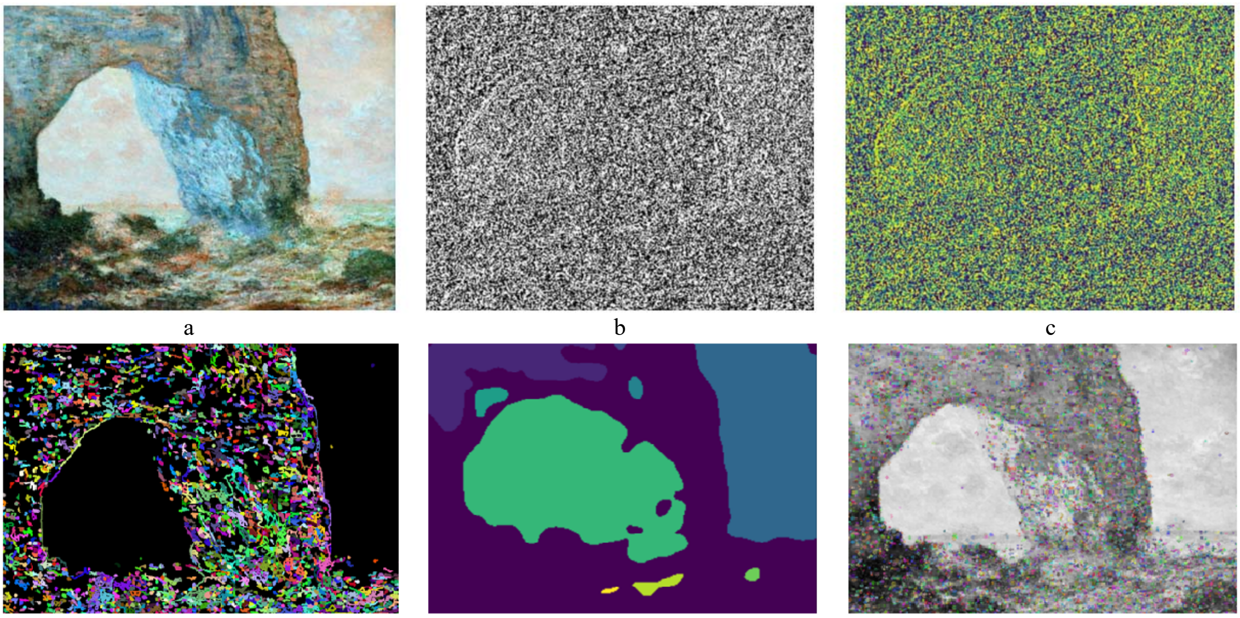
The paper considers the problem of automatic paintings classification by age. The authors propose a solution in the form of a classifier, which action is based on a weighted k-nearest neighbors algorithm.

To ensure high accuracy in the work, a set of attributes is proposed, including color, texture, statistical and other characteristics of images. Attribute values are generated from paintings photos in an intelligent decision-making system, which then classifies the painting by age.

To calculate the weights during k-nearest neighbors algorithm implementation, it is proposed to use Fisher score; to calculate the similarity measure, the authors propose to apply the Euclidean metric.

As a dataset for experimental research, it was proposed to use a set that includes works of famous world, European and Ukrainian artists, as well as metadata with artists' biographies, life period, and paintings genres description.

The scientific novelty of the work consists in the formation of a set of descriptors for paintings photos processing, which provides an accurate categorization of paintings by the time of creation.



a b c
d e f

Figure 1 – The results of features vectors calculating:
a – the original image of C. Monet "La Manneport" (1883); image processed using descriptors: b – LBP; c – color LBP; d – the first four moments; e – Haralik texture features; f) SIFT

DOI 10.15588/1607-3274-2022-2-8

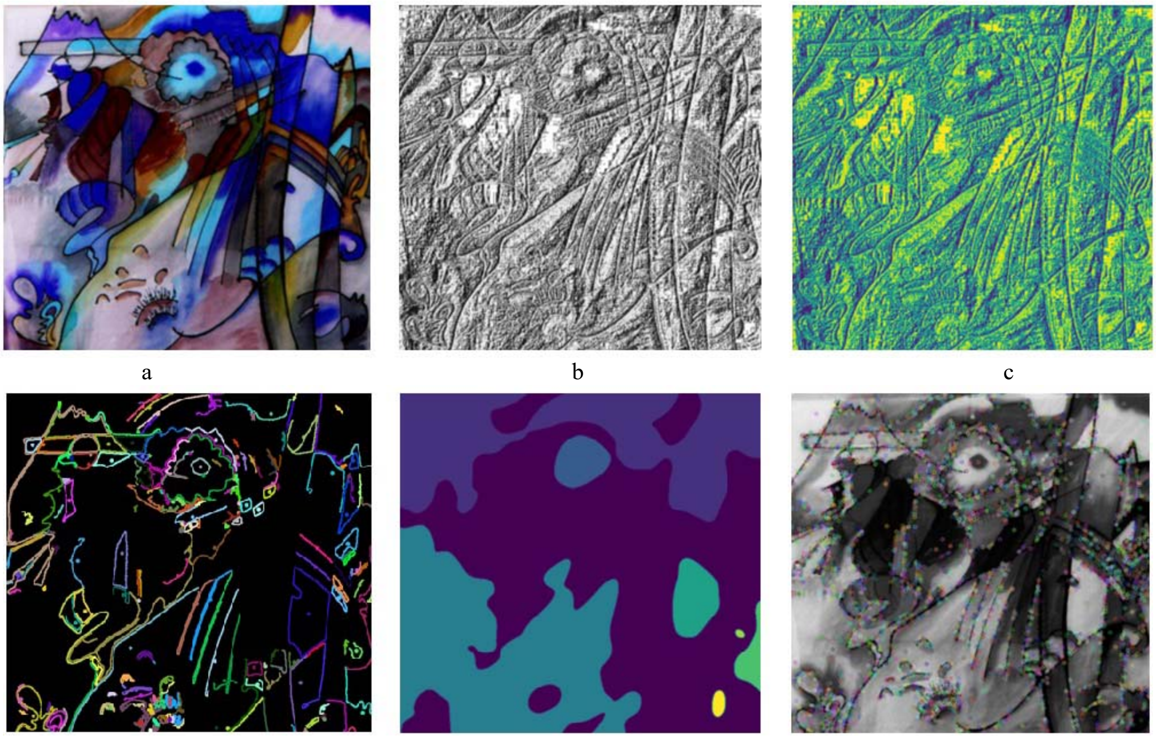


Figure 2 – The results of features vectors calculating:
a – the original image of V. Kandinsky "Composition VIII" (1923); image processed using descriptors: b – LBP; c – color LBP; d – the first four moments; e – Haralik texture features; f – SIFT

Table 2 – Accuracy of picture images classification depending on descriptor used to form the feature vector

Descriptor name	LBP	Color LBP	First 4 moments	Haralik texture features	Palette redundancy	SIFT descriptor	Tamura texture features	Generalized vector
Accuracy	70.92%	66.18%	62.81%	71.80%	63.89%	66.55%	72.92%	82.71%

Table 3 – Configuration parameters of classification algorithm

Algorithm configuration parameter name	Algorithm for constructing a search tree	Metric	Technique for data point weights calculating	Best number of neighboring points in a neighborhood	Best leaf size of a search tree
Parameter value	BallTree	Euclidean	Weight value is reciprocal of distance between the points	11	1

Table 4 – The results of paintings classification by artists by year of creation

Name	Amedeo Modigliani	Vasiliy Kandinskiy	Diego Rivera	Claude Monet	Rene Magritte	Salvador Dali	Gustav Klimt	Kazimir Malevich	Mikhail Vrubel
Accuracy	82.55%	80.48%	80.02%	82.29%	81.62%	81.92%	81.80%	80.71%	82.18%

The practical significance of the results is reducing the time and cost of customs verification of the paintings export legality.

Prospect for further research is related to further improving the accuracy of categorization by modifying the *k*-nearest neighbors algorithm.

ACKNOWLEDGEMENTS

The work was performed at the Department of Applied Mathematics, Kharkiv National University of Radio Electronics and the Department of Software Engineering, Dnipro University of Technology within the framework of scientific research conducted by the departments.

DOI 10.15588/1607-3274-2022-2-8

REFERENCES

1. Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property 1970 [Electronic resource]. Access mode: http://portal.unesco.org/en/ev.php-URL_ID=13039&URL_DO=DO_TOPIC&URL_SECTION=201.html
2. UNIDROIT Convention on Stolen or Illegally Exported Cultural Objects [Electronic resource]. Access mode: https://www.unidroit.org/instruments/cultural-property/1995-convention/
3. European Cultural Convention (ETS No. 018) [Electronic resource]. Access mode: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=018
4. Martynenko A., Moroz V., Hulina I. An intelligent decision support system for cultural property identification, *Computer-Integrated Technologies: Education, Science, Production*, 2020, Vol. 39, pp. 78–82. URL: http://cit-journal.com.ua/index.php/cit/article/view/126
5. Martynenko A.A., Tevyashev A. D., Kulishova N. Ye., Moroz B. I. System analysis of the problem of establishing the authenticity and authority of painting works, *System Research and Information Technologies,* 2022, No. 1, pp. 1–16.
6. Fiorucci M. Khoroshiltseva M., Pontil M., Traviglia A., Del Bue A., James S. Machine learning for cultural heritage: A survey, *Pattern Recognition Letters*, 2020, Vol. 133, pp. 102–108.
7. Belhi A., Bouras A., Al-Ali A. Kh., Sadka A. H., Eds. Data Analytics for Cultural Heritage: Current Trends and Concepts. – Springer, 2021, 279 p.
8. Shamir L., Macura T., Orlov N., Eckley D. M., Goldberg I. G. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art, *ACM Transactions on Applied Perception (TAP),* 2010, Vol. 7(2), pp. 1–17.
9. Falomir Z., Museros L., Sanz I., Gonzalez-Abril L. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qartlearn), *Expert Systems with Applications*, 2018, Vol. 97, pp. 83–94.
10. Sabatelli M., Kestemont M., Daelemans W., Geurts P. Deep transfer learning for art classification problems, *Proceedings of the European Conference on Computer Vision (ECCV),* 2018. https://doi.org/10.1007/978-3-030-11012-3_48.
11. Lee S.-G., Cha E.-Y. Style classification and visualization of art painting's genre using self-organizing maps, *Human-centric Computing and Information Sciences*, December 2016, Vol. 6(1). DOI: 10.1186/s13673-016-0063-4
12. Falomir Z., Museros L., Sanz I., Gonzalez-Abril L. Categorizing Paintings in Art Styles Based on Qualitative Color Descriptors, Quantitative Global Features and Machine Learning (QArt-Learn), *Expert Systems with Applications,* 2018, Vol. 97, pp. 83–94. doi.org/10.1016/j.eswa.2017.11.056
13. Jankovic´ R. Machine Learning Models for Cultural Heritage Image Classification: Comparison Based on Attribute Selection, *Information,* 2020, Vol. 11, 12. DOI: 10.3390/info11010012
14. Ye Z., Su L. The use of data mining and artificial intelligence technology in art colors and graph and images of computer vision under 6G internet of things communication, *International Journal of System Assurance Engineering and Management, The Society for Reliability, Engineering Quality and Operations Management (SREQOM).* India, and Division of Operation and Maintenance, Lulea University of Technology, Sweden, August, 2021. Vol. 12(4), pp. 689–695. DOI: 10.1007/s13198-021-01063-5
15. Wu Y. Application of Improved Boosting Algorithm for Art Image Classification, *Scientific Programming*, 2021, Vol. 2021. doi.org/10.1155/2021/3480414
16. Hamilton M., Fu S., Lu M., Bui J., Bopp D., Chen Zh., Tran F., Wang M., Rogers M., Zhang L., Hoder C., Freeman W. T. MosAIc: Finding Artistic Connections across Culture with Conditional Image Retrieval, *Proceedings of Machine Learning Research,* 2021, Vol. 1, pp. 1–23. arXiv:2007.07177v3
17. Pancaroglu D. Eds.: D.C. Wyld et al. Artist, style and year classification using face recognition and clustering with convolutional neural networks, *COMIT, SIPO, AISCA, MLIOB, BDHI,* 2020, CS & IT, CSCP 2020, pp. 41–54. DOI: 10.5121/csit.2020.101604
18. Banar N., Sabatelli M., Geurts P., Daelemans W., Kestemont M. Transfer Learning with Style Transfer between the Photorealistic and Artistic Domain, *Electronic Imaging, Computer Vision and Image Analysis of Art*, 2021, Vol. 41, pp. 1–9. doi.org/10.2352/ISSN.2470-1173.2021.14.CVAA-041
19. Cetinic E., Lipic T., Grgic S. Fine-tuning Convolutional Neural Networks for fine art classification, *Expert Systems with Applications*, 2018, Vol. 114, pp. 107–118. doi.org/10.1016/j.eswa.2018.07.026
20. Chen T., Yang J. A Novel Multi-Feature Fusion Method in Merging Information of Heterogenous-View Data for Oil Painting Image Feature Extraction and Recognition, *Front. Neurorobot.,* 2021, Vol. 12. doi.org/10.3389/fnbot.2021.709043
21. Zhao W., Zhou D., Qiu X., Jiang W. Compare the performance of the models in art classification, *PLoS ONE*, 2021, Vol. 16(3). doi.org/10.1371/journal.pone.0248414
22. Guo Z., Zhang L., Zhang D. A completed modeling of local binary pattern operator for texture classification, *IEEE Trans. on Image Processing,* 2010, Vol. 19, Issue 6, pp. 1657–1663.
23. Haralick R. M., Shanmugam K., Dinstein I. Textural Features for Image Classification, *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, Vol. 3, pp. 610–621. doi:10.1109/TSMC.1973.4309314
24. Otero I. R., Delbracio M. Anatomy of the SIFT method, *Image Processing On Line*, 2014, Vol. 4, pp. 370–396.
25. Wang D., Chen Y. Calculation and Application of Xin'an Painting School Art Style Model, *ICAITA 2020,* 2020, Vol. 1651, 012033. DOI: 10.1088/1742-6596/1651/1/012033
26. Bentley J. L. Multidimensional binary search trees used for associative searching, *Communications of the ACM,* 1975, Vol. 18 (9), pp. 509–517. DOI: 10.1145/361002.361007. S2CID 13091446.
27. Omohundro S. M. Five balltree construction algorithms, *International Computer Science Institute Technical Report,* 1989.
28. Best Artworks of All Time [Electronic resource]. – Access mode: https://www.kaggle.com/ikarus777/best-artworks-of-all-time/tasks

Received 24.01.2022.
Accepted 17.02.2022.

УДК 303.732.4 + 004.67 + 004.8

АВТОМАТИЧНА КЛАСИФІКАЦІЯ КАРТИН ЗА РОКОМ СТВОРЕННЯ

Мартиненко А. А. – старший викладач кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Тевяшев А. Д. – д-р техн. наук, професор, завідувач кафедри прикладної математики Харківського національного університету радіоелектроніки, м. Харків, Україна.

Кулішова Н. Є. – канд. техн. наук, доцент, професор кафедри медіасистем і технологій Харківського національного університету радіоелектроніки, Харків, Україна.

DOI 10.15588/1607-3274-2022-2-8

Мороз Б. І. – д-р техн. наук, професор, член-кореспондент Академії прикладної електроніки, професор кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», м. Дніпро, Україна.

Сергієнко О. С. – студентка кафедри прикладної математики Харківського національного університету радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглядається завдання автоматичної перевірки легітимності експорту творів живопису.

Мета. Запропоновано метод автоматичного визначення віку картини з цифрової фотографії за допомогою класифікації, яку виконує інтелектуальна система прийняття рішень.

Метод. Пропонується використовувати атрибут року створення картини як головний критерій для прийняття рішення під час митної перевірки легітимності експорту. Замість тривалої та дорогої музейної експертизи застосовується фотографування творів живопису в умовах митниці та обробка фото за допомогою набору дескрипторів. До набору дескрипторів пропонується включити локальні бінарні патерни, їх колірну модифікацію, текстурні ознаки Хараліка, перші чотири моменти, текстурні ознаки Тамури, SIFT дескриптор. Дані, отримані внаслідок дії дескрипторів, утворюють значення кількох десятків окремих атрибутів. Вони формують вектори даних, які потім конкатенуються в узагальнений опис вектора-об'єкта. У просторі ознак, створеному таким чином, виконується автоматична класифікація методом зважених *k*-найближчих сусідів. Пропонований алгоритм розраховує відстань між об'єктами в багатовимірному просторі значень атрибутів, і відносить нові об'єкти до сформованих класів. Критерієм для створення класів є вік картини із існуючої бази даних. Як міру близькості об'єктів пропонується використовувати метрики Евкліда та Мінковського. Розрахунок вагів для алгоритму класифікації запропоновано виконувати методом Фішера.

Результати. Ефективність запропонованого методу була досліджена під час експериментів із базою зображень, що містить фото картин світових, європейських та українських художників. Знайдено параметри конфігурації алгоритму, що забезпечують високу точність класифікації.

Висновки. Проведені експерименти показали ефективність вибраних дескрипторів формування векторів-описів зображень картин. Найбільшу точність забезпечує поєднання дескрипторів, яке виявляє суттєві відмінності у структурних властивостях зображень. Такий підхід до опису об'єктів у поєднанні із запропонованим алгоритмом класифікації та обраним головним критерієм забезпечує високу точність отриманих рішень. Напрямок подальших досліджень може включати використання згорткових нейронних мереж для підвищення точності класифікації за умови статичності бази даних.

КЛЮЧОВІ СЛОВА: інтелектуальна система прийняття рішень, автоматична класифікація, *k*-найближчих сусідів, дескриптори зображень, вектор ознак, митна експертиза, твори живопису.

УДК 303.732.4 + 004.67 + 004.8

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ КАРТИН ПО ГОДУ СОЗДАНИЯ

Мартыненко А. А. – старший преподаватель кафедры программного обеспечения компьютерных систем Национального технического университета «Днепровская политехника», Днепр, Украина.

Тевяшев А. Д. – д-р техн. наук, профессор, заведующий кафедрой прикладной математики Харьковского национального университета радиоэлектроники, Харьков, Украина.

Кулишова Н. Е. – канд. техн. наук, доцент, профессор кафедры медиасистем и технологий Харьковского национального университета радиоэлектроники, Харьков, Украина.

Мороз Б. И. – д-р техн. наук, профессор, член-корреспондент Академии прикладной электроники, профессор кафедры программного обеспечения компьютерных систем Национального технического университета «Днепровская политехника», Днепр, Украина.

Сергиенко А. С. – студентка кафедры прикладной математики Харьковского национального университета радиоэлектроники, Харьков, Украина.

АННОТАЦИЯ

Актуальность. Рассматривается задача автоматической проверки легитимности экспорта произведений живописи.

Цель. Предложен метод автоматического определения возраста картины по цифровой фотографии с помощью классификации, которую выполняет интеллектуальная система принятия решений.

Метод. Предлагается использовать атрибут года создания картины в качестве главного критерия для принятия решения в ходе таможенной проверки легитимности экспорта. Вместо длительной и дорогостоящей музейной экспертизы применяется фотографирование произведений живописи в условиях таможни и обработка фото с помощью набора дескрипторов. В набор дескрипторов предлагается включить локальные бинарные паттерны, их цветовую модификацию, текстурные признаки Харалика, первые четыре момента, текстурные признаки Тамуры, SIFT дескриптор. Данные, полученные в результате действия дескрипторов, образуют значения нескольких десятков частных атрибутов. Они формируют векторы данных, которые затем конкатенируются в обобщенный вектор-описание объекта. В пространстве признаков, созданном таким образом, выполняется автоматическая классификация методом взвешенных *k*-ближайших соседей. Предлагаемый алгоритм рассчитывает расстояние между объектами в многомерном пространстве значений атрибутов, и относит новые объекты к уже сформированным классам. Критерием для создания классов является возраст картины из существующей базы данных. В качестве меры близости объектов предлагается использовать метрики Евклида и Минковского. Расчет весов для алгоритма классификации предложено выполнять методом Фишера.

Результаты. Эффективность предложенного метода была исследована в ходе экспериментов с базой изображений, содержащей фото картин мировых, европейских и украинских художников. Найдены параметры конфигурации алгоритма, которые обеспечивают высокую точность классификации.

DOI 10.15588/1607-3274-2022-2-8

Выводы. Проведенные эксперименты показали эффективность выбранных дескрипторов для формирования векторов-описаний изображений картин. Наибольшую точность обеспечивает объединение дескрипторов, которое обнаруживает существенные различия в структурных свойствах изображений. Такой подход к описанию объектов в сочетании с предложенным алгоритмом классификации и выбранным главным критерием обеспечивает высокую точность полученных решений. Направление дальнейших исследований может включать использование сверточных нейронных сетей для повышения точности классификации при условии статичности базы данных.

КЛЮЧЕВЫЕ СЛОВА: интеллектуальная система принятия решений, автоматическая классификация, *k*-ближайших соседей, дескрипторы изображений, вектор признаков, таможенная экспертиза, произведения живописи.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property 1970 [Electronic resource]. – Access mode: http://portal.unesco.org/en/ev.php-URL_ID=13039&URL_DO=DO_TOPIC&URL_SECTION=201.html
2. UNIDROIT Convention on Stolen or Illegally Exported Cultural Objects [Electronic resource]. – Access mode: https://www.unidroit.org/instruments/cultural-property/1995-convention/
3. European Cultural Convention (ETS No. 018) [Electronic resource]. – Access mode: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=018
4. Martynenko A. An intelligent decision support system for cultural property identification / A. Martynenko, V. Moroz, I. Hulina // Computer-Integrated Technologies: Education, Science, Production. – 2020. – Vol. 39. – P. 78–82. URL: http://cit-journal.com.ua/index.php/cit/article/view/126
5. System analysis of the problem of establishing the authenticity and authority of painting works / [A. A. Martynenko, A. D. Tevyashev, N. Ye. Kulishova, B. I. Moroz] // System Research and Information Technologies. – 2022. – № 1. – P. 1–16.
6. Fiorucci M. Machine learning for cultural heritage: A survey / [M. Fiorucci, M. Khoroshiltseva, M. Pontil et al.] // Pattern Recognition Letters. – 2020. – Vol. 133. – P. 102–108.
7. Data Analytics for Cultural Heritage: Current Trends and Concepts / Eds.: A. Belhi, A. Bouras, A. Kh. Al-Ali, A. H. Sadka. – Springer, 2021. – 279 p.
8. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art / [L. Shamir, T. Macura, N. Orlov et al.] // ACM Transactions on Applied Perception (TAP). – 2010. – Vol. 7(2). – P. 1–17.
9. Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qartlearn) / [Z. Falomir, L. Museros, I. Sanz, L. Gonzalez-Abril] // Expert Systems with Applications. – 2018. – Vol. 97. – P. 83–94.
10. Deep transfer learning for art classification problems / [M. Sabatelli, M. Kestemont, W. Daelemans, P. Geurts] // Proceedings of the European Conference on Computer Vision (ECCV), 2018. https://doi.org/10.1007/978-3-030-11012-3_48.
11. Lee S.-G. Style classification and visualization of art painting's genre using self-organizing maps / S.-G. Lee, E.-Y. Cha // Human-centric Computing and Information Sciences, December. – 2016. – Vol. 6(1). DOI: 10.1186/s13673-016-0063-4
12. Categorizing Paintings in Art Styles Based on Qualitative Color Descriptors, Quantitative Global Features and Machine Learning (QArt-Learn) / [Z. Falomir, L. Museros, I. Sanz, L. Gonzalez-Abril] // Expert Systems with Applications. – 2018. – Vol. 97. – P. 83–94. doi.org/10.1016/j.eswa.2017.11.056
13. Jankovic´ R. Machine Learning Models for Cultural Heritage Image Classification: Comparison Based on Attribute Selection / R. Jankovic´ // Information. – 2020. – Vol. 11, 12. DOI: 10.3390/info11010012
14. Ye Z. The use of data mining and artificial intelligence technology in art colors and graph and images of computer vision under 6G internet of things communication / Z. Ye, L. Su // International Journal of System Assurance Engineering and Management, The Society for Reliability, Engineering Quality and Operations Management (SREQOM), India, and Division of Operation and Maintenance, Lulea University of Technology, Sweden, August. – 2021. – Vol. 12(4). – P. 689–695. DOI: 10.1007/s13198-021-01063-5
15. Wu Y. Application of Improved Boosting Algorithm for Art Image Classification/ Y. Wu // Scientific Programming, 2021. – Vol. 2021. doi.org/10.1155/2021/3480414
16. Hamilton M. MosAIc: Finding Artistic Connections across Culture with Conditional Image Retrieval / [M. Hamilton, S. Fu, M. Lu et al.] // Proceedings of Machine Learning Research. – 2021. – Vol. 1. – P. 1–23. arXiv:2007.07177v3
17. Pancaroglu D. Artist, style and year classification using face recognition and clustering with convolutional neural networks / D. Pancaroglu // Eds.: D. C. Wyld et al. COMIT, SIPO, AISCA, MLIOB, BDHI – 2020, CS & IT – CSCP 2020. – P. 41–54. DOI: 10.5121/csit.2020.101604
18. Transfer Learning with Style Transfer between the Photorealistic and Artistic Domain / [N. Banar, M. Sabatelli, P. Geurts et al] // Electronic Imaging, Computer Vision and Image Analysis of Art. – 2021. – Vol. 41. – P. 1–9. doi.org/10.2352/ISSN.2470-1173.2021.14.CVAA-041
19. Cetinic E. Fine-tuning Convolutional Neural Networks for fine art classification / E. Cetinic, T. Lipic, S. Grgic // Expert Systems with Applications. – 2018. – Vol. 114. – P. 107–118. doi.org/10.1016/j.eswa.2018.07.026
20. Chen T. A Novel Multi-Feature Fusion Method in Merging Information of Heterogenous-View Data for Oil Painting Image Feature Extraction and Recognition / T. Chen, J. Yang // Front. Neurorobot. – 2021. – Vol. 12. doi.org/10.3389/fnbot.2021.709043
21. Compare the performance of the models in art classification / [W. Zhao, D. Zhou, X. Qiu, W. Jiang] // PLoS ONE. – 2021. – Vol. 16(3). doi.org/10.1371/journal.pone.0248414
22. Guo Z. A completed modeling of local binary pattern operator for texture classification / Z. Guo, L. Zhang, D. Zhang // IEEE Trans. on Image Processing. – 2010. – Vol. 19, Issue 6. – P. 1657–1663.
23. Haralick, R. M. Textural Features for Image Classification / R. M. Haralick, K. Shanmugam, I. Dinstein // IEEE Transactions on Systems, Man, and Cybernetics. – 1973. – Vol. 3. – P. 610–621. DOI: 10.1109/TSMC.1973.4309314
24. Otero I. R. Anatomy of the SIFT method / I. R. Otero, M. Delbracio // Image Processing On Line. – 2014. – Vol. 4. – P. 370–396.
25. Wang D. Calculation and Application of Xin'an Painting School Art Style Model / D. Wang, Y. Chen // ICAITA 2020. – 2020. – Vol. 1651. – 012033. DOI: 10.1088/1742-6596/1651/1/012033
26. Bentley J. L. Multidimensional binary search trees used for associative searching / J. L. Bentley // Communications of the ACM. – 1975. – Vol. 18 (9). – P. 509–517. doi:10.1145/361002.361007. S2CID 13091446.
27. Omohundro S. M. Five balltree construction algorithms / S. M. Omohundro // International Computer Science Institute Technical Report, 1989.
28. Best Artworks of All Time [Electronic resource]. – Access mode: https://www.kaggle.com/ikarus777/best-artworks-of-all-time/tasks

DOI 10.15588/1607-3274-2022-2-8