

# Correlative features for the classification of textural images

M A Turkova<sup>1</sup> and A V Gaidel<sup>1,2</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute - Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Molodogvardeyskaya str. 151, Samara, Russia, 443001

**Abstract.** The work is aimed to study various correlative image features in order to solve the classification problem of textural images, with further selection of the most informative features for increasing the reliability of classification. In the present paper series of experiments are performed on a set of 4480 real digital images from the open Kylberg Texture Dataset v. 1.0. A set of programs necessary for computations was developed. The autocorrelative image features are calculated, and the approximating coefficients from the regression equation are obtained. Due to the removal of the least informative features, the share of correctly recognized images increased in 1.04 times and amounted to 90.11% for 28 image classes.

## 1. Introduction

At present, the task of computer vision and, consequently image recognition, cause a particular interest in many areas of science and technology. Identifying a number of characteristic features which help to classify a texture is essential for the textural images recognition. A variety of parameters can act as the image features. Optimal and universal descriptors development continues. Among the most common classification signs for textural images are morphological features, fractal dimension, Markov's parameters, samples of the autocorrelation function and others [1].

At the beginning of the classification process each image can be characterized by a large number of parameters. Since the classifier for solving the classification problem is not ideal, the error it introduces essentially depends on the quality of the feature space. Thus, for proper classification, it is necessary to reduce the subset of features by removing uninformative and redundant features, thereby reducing the dimensionality of the feature set. Usage of the optimal feature subset will increase the reliability, at the same time reducing the computational complexity and time for classification.

Despite the urgency of the data processing problem, an optimal method for identifying and selecting features has not been found yet. In paper [2] a novel feature selection strategy resistant to problems with asymmetric data and based on Genetic Programming is proposed. It works well with both balanced and unbalanced data. The results of the experiments made possible reducing the size of the feature space up to 83%, which allow one to increase the effectiveness of the learning process. More details about the process of selecting features and using the samples of the correlation function as characteristics are described in [3].

A two-stage feature selection approach which makes full use of interactions is proposed in [4]. On the first stage, the problem of feature selection is decomposed into the sum of information interaction. Then, higher-order interactions are used to select interaction-preserving feature subset. On the second stage, the design of experiments is employed to identify significant interactions from the subset of

features. A flexible algorithm for selecting a subset of features under user control based on clustering was developed in work [5]. This is a two-step hybrid process which reduces the computational complexity of the feature selection, especially for large sets of features.

In [6] the feature selection is carried out in two stages: in the first step, the correspondence of each objective function feature is evaluated. If the attribute is irrelevant, then it is discarded. In the second step, the number of features is minimized by removing duplicate or low-performance functions. In [7] the process of selecting features is carried out using a heuristic evolutionary algorithm, which is the example of competition in a tribe. The main idea of the method is that only the most significant signs win and participate in the classification.

In this paper, there is a data set consisting of different classes of textural images. A number of features is formed for each image, based on the readings of the correlation function and the coefficients from the regression equations. A filtering method with a marked set of training data is used for the feature selection. A measure of the Euclidean distance is chosen as a search strategy.

## 2. Texture descriptors

### 2.1. Autocorrelation function as a characteristic of a texture image

To form a set of texture image features we use the often applicable texture descriptor – the autocorrelation function, as in [8]:

$$\rho(x, y) = \frac{1}{(N_i - |x|)(N_j - |y|)} \sum_i \sum_j I(i, j) I(i + x, j + y) \left( \frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} I(i, j)^2 \right)^{-1} \quad (1)$$

where  $I(i, j)$  – pixel value,  $N_i \times N_j$  – size of image,  $\rho(0,0) \equiv 1$  [1]. It was shown in [9] that linear combinations of autocorrelation function counts are the best signs in the class of quadratic features.

The coefficients of the autocorrelation function determine the relationship of the pixels in the image. As the distance between points increases, the correlation weakens and the value of the coefficients does not make a significant contribution to the calculations. Before the study begins, it is necessary to normalize the autocorrelation function to get rid of the high-frequency noise influence. Thus, we obtain an expression for the coefficients of the autocorrelation function.

### 2.2. Regression equation coefficients as a characteristic of the texture image

The autocorrelation function  $\rho(x, y)$  obtained in the previous section can be unnecessarily complicated for describing the texture. However, if the existing function is approximated by a parametric function, then its parameters can be used as a more reliable texture descriptor.

Suppose that the autocorrelation function at a point  $(x_i, y_i)$  is known and is equal to  $\rho_i$ . Let us approximate a function by the following regression equations:

$$\rho_1(x, y) = \exp(a_1 x^2 + a_2 y^2 + a_3), \quad (2)$$

$$\rho_2(x, y) = \exp\left(b_1 \sqrt{x^2 + y^2} + b_2\right), \quad (3)$$

where  $a_1, a_2, a_3, b_1, b_2$  – the coefficients to be found. We take the logarithm of (2) and (3) and obtain:

$$\ln(\rho_1(x, y)) = a_1 x^2 + a_2 y^2 + a_3, \quad (4)$$

$$\ln(\rho_2(x, y)) = b_1 \sqrt{x^2 + y^2} + b_2. \quad (5)$$

The total error for (4) and (5), respectively, is:

$$E_1 = \sum_i (\ln(\rho_1(x_i, y_i)) - \rho_i)^2 = \sum_i (a_1 x_i^2 + a_2 y_i^2 + a_3 - \rho_i)^2,$$

$$E_2 = \sum_i (\ln(\rho_2(x_i, y_i)) - \rho_i)^2 = \sum_i (b_1 \sqrt{x_i^2 + y_i^2} + b_2 - \rho_i)^2.$$

It is necessary to select the parameters  $a_1, a_2, a_3, b_1, b_2$  in such a way for the error to be minimal. To do this we take the partial derivative with respect to each parameter and equate it to zero, obtaining a system of linear equations with respect to unknown parameters:

$$\begin{cases} \frac{\partial E_1}{\partial a_1} = 0 \Rightarrow 2 \sum_i (a_1 x_i^2 + a_2 y_i^2 + a_3 - \rho_i) x_i^2 = 0, \\ \frac{\partial E_1}{\partial a_2} = 0 \Rightarrow 2 \sum_i (a_1 x_i^2 + a_2 y_i^2 + a_3 - \rho_i) y_i^2 = 0, \\ \frac{\partial E_1}{\partial a_3} = 0 \Rightarrow 2 \sum_i (a_1 x_i^2 + a_2 y_i^2 + a_3 - \rho_i) = 0. \end{cases}$$

Solving the resulting system of linear equations:

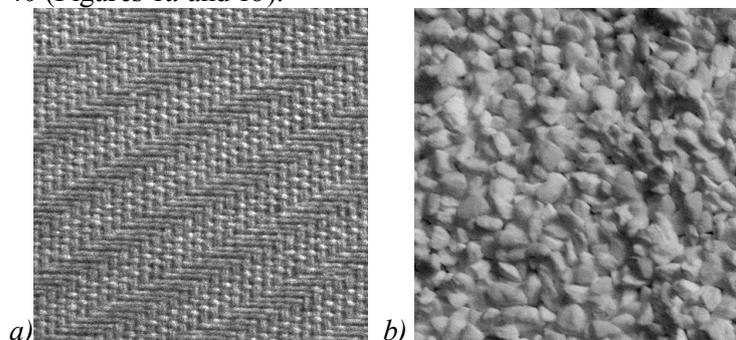
$$\begin{cases} a_1 \sum_i x_i^4 + a_2 \sum_i x_i^2 y_i^2 + a_3 \sum_i x_i^2 = \sum_i \rho_i x_i^2, \\ a_1 \sum_i x_i^2 y_i^2 + a_2 \sum_i y_i^4 + a_3 \sum_i y_i^2 = \sum_i \rho_i y_i^2, \\ \sum_i x_i^2 + a_2 \sum_i y_i^2 + a_3 \sum_i 1 = \sum_i \rho_i, \end{cases}$$

we obtain expressions for the coefficients  $a_1$ ,  $a_2$ ,  $a_3$  that are used in the regression equation (2). Similarly, we obtain values for the coefficients  $b_1$ ,  $b_2$  from equation (3) [1].

### 3. Experimental studies

#### 3.1. Input data

As the input data set 28 classes of textural images from the open image database Kylberg Texture Dataset v. 1.0 are used [10]. Each class contains 160 unique monochrome photographs without rotations measuring 576×576 pixels. All images are normalized with a mean value of 127 and a standard deviation of 40 (Figures 1a and 1b).

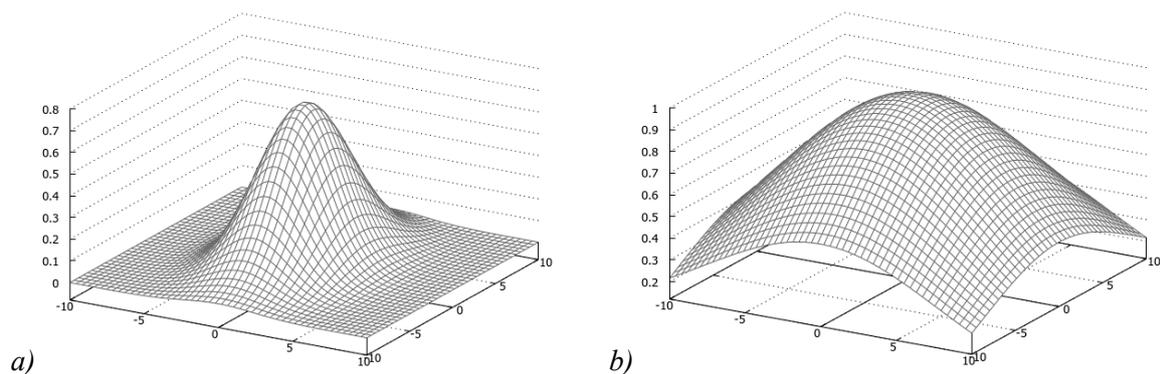


**Figure 1.** Examples of images on which experiments were performed: seat (a), pearl sugar (b).

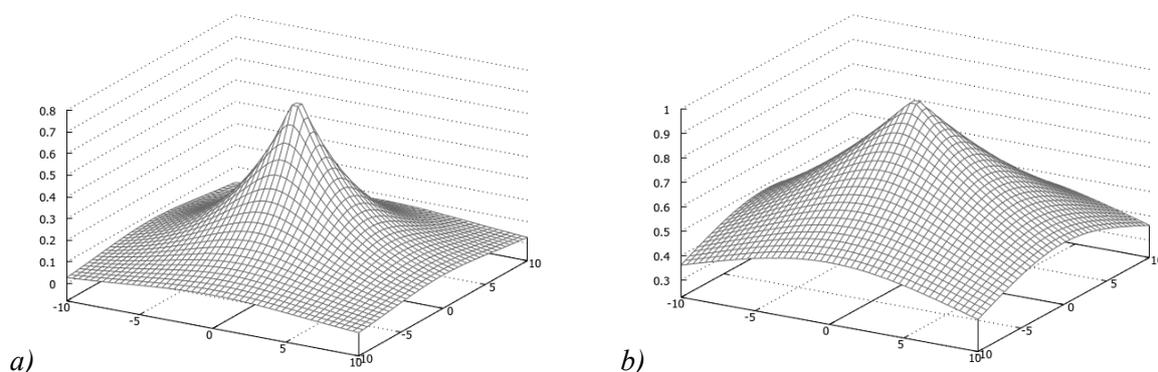
For each image a vector of 30 features is constructed. The first 25 features are formed as the samples of the correlation function  $\rho(x, y)$  (1),  $x = \overline{0, 4}$ ,  $y = \overline{0, 4}$ , but instead of the value  $\rho(0, 0) \equiv 1$  the count  $\rho(-1, 1)$  is used. Features 25-27 are the coefficients  $a_1$ ,  $a_2$ ,  $a_3$  from the regression equation (2), 28 and 29 are the coefficients  $b_1$ ,  $b_2$  from equation (3). The form of the regression equations (2) and (3) for the specific values obtained during the computational experiment is shown in Figures 2a-b and 3a-b, respectively. The values of the coefficients from the regression equations (2) and (3) are presented in table 1.

**Table 1.** The values of the coefficients of the regression equations for the graphs presented in Figures 1a and 1b.

Coefficients	$a_1$	$a_2$	$a_3$	$b_1$	$b_2$
Seat	-0.08999	-0.03208	-0.30889	-0.23781	-0.23272
Pearl sugar	-0.00835	-0.00652	-0.04151	-0.02968	-0.02982



**Figure 2.** The graph of the regression equations (2): for Figure 1a (a), for Figure 1b (b)



**Figure 3.** The graph of the regression equations (3): for Figure 1a (a), for Figure 1b (b)

Also, the "Class" field is added to the characteristics vector, which has a range of values from 1 to 28 and will be used for the model training.

### 3.2. The scheme of the computational experiment

Before the beginning of the research it is necessary to scale out the feature set, since the numerical values are presented in different scales, according to the following formulas. To normalize the sample it is necessary to divide each feature  $x_i$  into the sample rate:

$$x_i = x_i \left( \sum_{k=1}^K x_k \right)^{-1/2}, i = \overline{1, K},$$

where  $K$  is the total number of features. For standardization, we use the formula:

$$x_i = (x_i - M(i))R(i, i)^{-1/2}, i = \overline{1, K},$$

where  $M(i)$  is the mean value and  $R(i, i)$  is the standard deviation.

To classify images, it is necessary to form a training sample. In this paper we form it according to the principle of leave-one-out cross-validation: at each iteration step of the training there is a sample of  $K - 1$  elements, where  $K$  is the total sample size. Due to this, we increase the accuracy of classification (it is defined as the ratio of the number of correctly defined classes to their total number), although we increase the computational complexity of the algorithm. We distribute an image with an unknown class to a particular group based on the nearest neighbour, finding the vector closest to the desired vector by the set of characteristics.

During the first computational experiment we form the training sample consisting of the object by the cross-validation method. We successively determine the class of each image not included in the training sample. As a feature set, we use all the 30 features in the first case, in the second – only features with numbers 0-24, in the third – 25-29. During the second computational experiment, the training sample is formed in a similar way, but the attribute space is formed by the full search method

as a subset of the cardinalities one and three, as well as the random search method with capacities of four and five.

To compare the results of computational experiments we will classify the vectors for each image and evaluate the quality criterion  $J$ , assigned as the proportion of correctly recognized objects to their total number. Object classification should be carried out in its characteristic space, obtained according to the requirements of the computational experiments described above.

### 3.3. Results of experiments

The obtained results for the classification of texture images in accordance with the task of the first computational experiment are presented in table 2.

**Table 2.** Results of the first computational experiment.

Number of features	Cardinality	Reliability, %	Time, s
0, 1, ..., 29	30	86.93	876.71
0, 1, ..., 24	25	69.82	732.74
25, 26, 27, 28, 29	5	53.19	72.59

Table 2 shows that the highest proportion of correctly classified objects is achieved using a full set of features. However, in this case, the presence of non-informative features is possible, which will reduce the reliability of the classification. Also, a large subset of characteristics significantly increases the computational complexity of the algorithm and, as a consequence, increases the calculation time. Therefore, it is necessary to select features, which will reduce re-training, increase reliability and shorten the classification time.

The most informative features for the classification problem were signs under numbers 4, 0 and 5. The individual qualities of the five best attributes are given in table 3.

**Table 3.** Individual qualities of features.

Number of features	Reliability, %
4	44.02
0	35.09
5	32.54
9	30.84
24	30.22

We form feature subsets of cardinality three by the method of complete enumeration. The average time necessary for classifying the images using the generated subsets was 72.36 seconds. Table 4 shows that the first two best subsets of the features consist of only the samples of the correlation function, the third one also includes a sign derived from the coefficient of the regression equation. The reliability of the classification using the most effective feature subset of cardinality three is 1.13 times lower than the classification using the full feature set. However, the time spent on classification is 11.38 times lower.

**Table 4.** The feature selection result using a subset of cardinality three.

Number of features	Reliability, %
4, 14, 19	77.05
3, 4, 19	76.38
4, 19, 29	76.38

We use the method of random search to construct subsets of cardinalities four and five. The classification using a feature subset of cardinality four is 1.05 times as low as the classification proposed in the first computational experiment. However, when using a subset of cardinality five reliability of classification is 1.04 times higher.

Table 5 shows that in the formation of a characteristic subset the most informative features also participate: in the subset of cardinality four are the sign at number four, in the subset of cardinality of five, the sign with the number zero and four. The average classification time was 67.14 and 73.44 seconds when using subsets of cardinalities four and five, which gives a gain in time of 13.06 and 11.94 times, respectively.

**Table 5.** The feature selection results using subsets of cardinalities four and five.

Number of features	Reliability, %
4, 14, 28, 29	82.77
4, 14, 19, 25	82.59
4, 19, 26, 29	82.59
0, 2, 3, 25, 29	90.11
0, 2, 3, 25, 27	89.69
0, 2, 3, 4, 29	89.22

Thus, after carrying out series of experiments, the usage of the feature subset of cardinality five was the most effective, with the help of which it was possible to achieve a quality classification of more than 90%. Also the need to reduce feature set by removing the least informative ones was confirmed experimentally.

#### 4. Conclusion

In the recent work, the objects were classified on the base of the calculated features. A number of features were formed using the samples of the correlation function and the coefficients from the regression equations. The selection of features was carried out with the methods of complete and random search.

The features that are the most informative and necessary in the construction of the classifier were revealed, they turned out to be the readings of the correlation function  $\rho(1, 0)$ ,  $\rho(0, 1)$ ,  $\rho(1, 1)$ . Their inclusion in a variety of attributes for further classification will increase its reliability, allowing one to remove less informative features. The features, which are coefficients from the regression equations, were not among the top ten, which indicates their little information in the task of classifying images. Among the features that have the least impact on the classification process, are the coefficients  $a_3$  and  $a_2$  from the regression equations  $\rho_1(x, y) = \exp(a_1x^2 + a_2y^2 + a_3)$  and  $\rho_2(x, y) = \exp(b_1\sqrt{x^2 + y^2} + b_2)$ , as well as the correlation function  $\rho(0,2)$ .

The highest reliability of the image classification, which amounted to 90.11% was achieved with the use of a feature subset of cardinality five, which included the samples of the correlation function  $\rho(0,1)$ ,  $\rho(0,3)$ ,  $\rho(0,4)$  and the coefficients  $a_1$  and  $b_2$  from the regression equations.

In this way, the hypothesis about the need to exclude the least informative features from the set was experimentally confirmed. When using the full set of attributes for image classification, the achieved reliability was 86.83%, which is 1.04 times as low as the maximum reliability obtained in the course of the computational experiment. Also, the time spent on classification decreased by 11.94 times by reducing the cardinality of feature set.

#### 5. References

- [1] Petrou M 2006 *Image processing: Dealing with texture* (John Wiley & Sons, Ltd)
- [2] Viegas F, Rocha L, Goncalves M, Mourao F, Salles T, Andrade G and Sandin I 2018 A Genetic Programming approach for feature selection in highly dimensional skewed *Neurocomputing* **273** 554-569
- [3] Gaidel A V, Zelter P M, Kapishnikov A V and Khramov A G 2014 Computed tomography texture analysis capabilities in diagnosing a chronic obstructive pulmonary disease *Computer Optics* **38(4)** 843-850
- [4] Tang X, Dai Y, Sun P and Meng S 2018 Interaction-based feature selection using Factorial Design *Neurocomputing* **281** 47-54

- [5] Goswami S, Das A K, Chakrabarti A and Chakraborty B 2017 A feature cluster taxonomy based feature selection technique *Expert System with Applications* **79** 76-89
- [6] Rahmaninia M and Moradi P 2017 OSFSMI: Online Stream Feature Selection Method based on Mutual Information *Appl. Soft Comput.* **68** 733-746
- [7] Maa B and Xia Y 2017 A tribe competition-based genetic algorithm for feature selection in pattern classification *Appl. Soft Comput.* **58** 328-338
- [8] Gaidel A V 2015 A method for adjusting directed texture features in biomedical image analysis problems *Computer Optics* **39(2)** 287-293 DOI: 10.18287/0134-2452-2015-39-2-287-293
- [9] Gaidel A V 2016 Matched polynomial features for the analysis of grayscale biomedical images *Computer Optics* **40(2)** 232-239 DOI: 10.18287/2412-6179-2016-40-2-232-239
- [10] Kylberg G 2011 The Kylberg Texture Dataset v. 1.0: External report (Blue series) *Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University* **35**

### **Acknowledgments**

The work was partially funded by the Russian Foundation of Basic Research grant 16-41-630761 p\_a, by the Federal Agency for Scientific Organizations under agreement No. 007-GZ/Ch3363/26 and by the Russian Federation Ministry of Education and Science.