

DEVELOPMENT OF COMPUTATIONAL PROCEDURE OF LOCAL IMAGE PROCESSING, BASED ON THE USAGE OF HIERARCHICAL REGRESSION

V.N. Kopenkov
Samara University
443068, Samara, Russia
vkop@geosamara.ru

ABSTRACT

The article deal with technology of digital images processing on the base of non-linear algorithms. This approach allows to construct the efficient procedure of local image processing. The aim of this research is to workout an algorithm of processing images with predetermined computational complexity and the best quality of processing on the existing data set avoiding a problem of retraining or lesstraining. To achieve this aim we use local discrete wavelet transformation and hierarchical regression to construct local image processing procedure on the base of a training dataset. Moreover, we workout method to estimate the necessity of finishing or continuing the training process. This method is based on of the functional of full cross-validation control which allow to construct processing procedure with predetermined computational complexity and veracity, and with the best quality.

Keywords

local processing, hierarchical regression, interval estimate, function of complete sliding quality control.

1. INTRODUCTION

The tasks of image processing and signal analysis need to be solved in different fields of human activity [Soi09, Woo05, Pra07]. Local processing of digital images is one of the most important kinds of transformation in the theory and practice of digital image processing and computer vision.

Historically, the first processing procedures used local linear methods that allow the construction of optimal (in some sense), processing procedures [Soi09, Pra07]. However, the taking into consideration of new digital signal processing tasks (processing of video, audio, satellite images, etc.), problems of processing large amounts of information (satellite images, remote sensing data, hyperspectral data, multi-dimensional signals), processing in real time, and needs of rising of processing efficiency resulted in the necessity of using nonlinear type of transformations [Soi09, Hai06]. One of the most common approach currently in use is the implementation of the cybernetic principle of the "black box" (the terms of other authors is the processing via recognition, processing basing on precedents and so on). In this case transformation itself and its parameters are determined by the

analyzing of the input and output signals, or images.

The classic approach to construction of approximately universal procedures of local adaptive digital signal and images processing, which implements the principle of "black box" is based on usage of artificial neural networks technique [Hai06]. An alternative, but substantially less researched version of described task solution based on using of a hierarchical computational structure, such as the decisions tree and regressions tree. [Bre84, Kop12]. This paper develops the idea of the creation of universal mechanism for the construction of local non-linear computational processing procedures based on a hierarchical scheme and features based on local discrete wavelet decomposition of the image.

In addition, the methodology of decision-making on stopping the learning process, as well as the veracity of the obtained results are also presented in the article. Usually, for estimation of the generalization capability and selection a stop learning rule for processing procedure the Vapnik–Chervonenkis statistical theory is used [Vap74]. This theory interconnects the three parameters of training: training error, veracity (reliability), and the length of training dataset. But estimates of statistical theory are highly overestimated and ignore the potential rearrangement of training and testing dataset elements. A more efficient way to estimate the generalization capability is the use of Vorontsov combinatorial theory [Vor04], which is based on evaluation of the functional of full cross-validation, assuming verification all possible combinations of division dataset into training and testing parts. The

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

correct solution to the task of construction of an image processing procedures which takes into account all combination of sets of training and control data is unrealizable in practice because of the giant search of variants of different combinations.

The paper is organized as follows. The first section devoted to subject introduction. The task specification and description of proposed solution, as well as a scheme of processing process are presented in the second section. Description and structure of the algorithm of construction of local image processing procedure and its parameters is presented in the third section. The fourth section is devoted to the description of methodology which allows to determine the rule for stop the process of formation and busting various combinations of training and control samples and stop construction process in

general. And finally, conclusions, recommendations, acknowledgements, and references are presented in the end of the paper.

2. TASK SPECIFICATION

2.1. Image local processing model

Model of the local image processing technology, which implement the principle of "black box" (processing through the recognition or based on precedents), suggests decomposition of the transformation in two stages: the formation of the image fragment description (local features computing) and calculation of transformation results. The general scheme of image processing is shown in Fig 1.

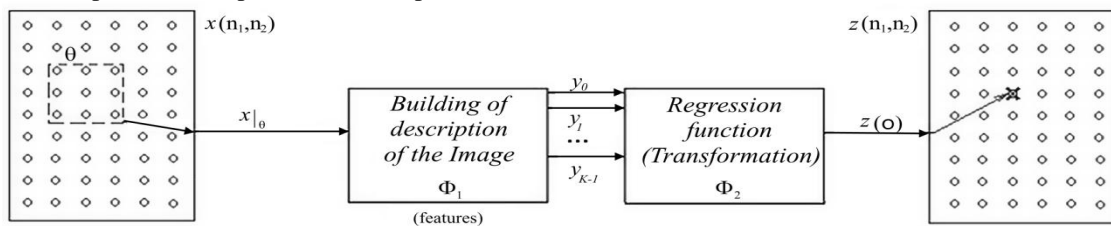


Figure 1. Local image processing scheme.

The main task in the first stage is the formation features (some specific set of image properties) for predetermined local image fragment – $\bar{y} = (y_0, y_1, \dots, y_{k-1})^T$, $\bar{y} \in \mathbf{R}^K$ on the base of transformation $\Phi_1 : \mathbf{R}^{M_1 \times M_2} \rightarrow \mathbf{R}^K$.

These features are used to calculate the result of transformation $\Phi_2 : \mathbf{R}^K \rightarrow \mathbf{K}$ (and to generate the resulting image Z) during the second stage of processing.

The whole construction process is based on the processing precedents – a set of matched pairs of images $\{x|_{\theta(n_1, n_2)}, z(n_1, n_2)\}_{(n_1, n_2) \in \Theta}$ (which are usually called *training dataset*) in order to minimize the processing error:

$$\varepsilon = \frac{1}{|\Theta|} \sum_{(n_1, n_2)} \|z - \Phi_2(\Phi_1(x))\| \rightarrow \min_{\Phi_1, \Phi_2}, \quad (1)$$

where Θ – the image domain, $\theta(n_1, n_2) \subseteq \Theta$ – restriction to the local fragment size $M_1 \times M_2$: $\theta(n_1, n_2) = \{(n_1 + m_1, n_2 + m_2) : m_1 = \overline{0, M_1 - 1}, m_2 = \overline{0, M_2 - 1}\}$.

2.2. Characteristics of the decision

The most known solution of described task is based on the usage of artificial neural networks. Such approach has some special features, advantages and disadvantages which are well described in detail [Hai06]. The alternative technology of the processing

procedure construction is based on special hierarchical computational structures, such as *regression trees* and *decision trees* [Bre84, Kop12]. These trees are the hierarchical structures consisting of 2 types of vertices - non-terminal vertices which define a partition of features domain, and terminal vertices which store a regression function.

The procedure based on regression trees has some advantages in comparison with the neural networks:

- automatic correction of "architecture" of the transformation;
- automatic selection of local features which is result of the partition process;
- finitely of the building and tuning process (computational efficiency);
- ease of tuning of the regression parameters in the terminal vertex.

There are some restrictions for the practical implementation. At first, the most important task on the stage of image features calculation in a "sliding window" mode is the task of development of a computationally efficient algorithm for this calculation. Moreover, this algorithm should allow consistently increase the features number up to whole system, because the traditional algorithms of the defining of the effective features subset based on iterative methods and computationally inefficient. At second, the main task on the stage of designing of hierarchical regression is the development of an algorithm to automatic construction of processing procedures on the base of the training dataset which

be able to avoid the retraining and insufficient training problems.

2.3. Choosing of the features types

We used the family of signal characteristics on the base of local wavelet discrete transformations (DWT) of the signals and images as an image features set. Such features have the following characteristics:

- ✓ existence of the computationally efficient calculation algorithm [Kop08];
- ✓ complete description of the input signal;
- ✓ consistent obtaining and usage of features removes the problem of iterates on the features set.

Issues related to the features formation on the base of local DWT algorithms, as well as their advantages and specialty in relation to local image processing tasks, are considered in the work [Kop08].

The classic scheme for fast calculating of local wavelet transformation (FWT) is based on Mallat scheme [Hai06] and in accordance to the theory of a multiple-scale analysis can be represented as following equations:

$$w_{l+1}^+(p) = \sum_{n \in D_h} h(n-2p) \cdot w_l^+(n),$$

$$w_{l+1}^-(p) = \sum_{n \in D_g} g(n-2p) \cdot w_l^+(n),$$

where $p = \overline{1, N}$, N – length of the input signal, $h(n), g(n)$ – such filters that: $\sum_n h(n) = 1$, $g(n) = (-1)^n h(-n+2t-1)$, ($t \in Z$), D_h, D_g – length of the filters, $l = \overline{0, \log_2 M}$ – wavelets levels, M – processing window size.

Concerning the image processing, the computational complexity of such algorithm for the wavelet levels $[L_1, L_2]$ can be evaluated [Kop08] as following:

FWT on the base of Mallat scheme:

$$U_1^*(L_1, L_2) = 8/3(2^{2L_2} - 1);$$

Modified FWT: $U_2^*(L_1, L_2) \underset{N \rightarrow \infty}{\approx} 8L_2 - 5L_1 - 5;$

Recursive FWT: $U_3^*(L_1, L_2) = 13(L_2 - L_1 + 1).$

3. ALGORITHM FOR CONSTRUCTION OF THE LOCAL IMAGE PROCESSING PROCEDURE

3.1. Regression tree construction

Technology of regression trees construction consists of the following stages:

1) Determination of parameters and features of hierarchical structure building process.

Necessary to select the vertices for further separation based on estimation error in them. Then determine the *parameters of vertices partitioning* (threshold and number of vertices to divide) and choose the "best feature" for partition. Such choice has to be done

taking into account that the aggregation of "best feature" and parameters of the partition should provide maximum errors reduction.

In the case of linear regression:

$$f_j(\bar{y}) = \sum_{k=0}^{K-1} a_{jk} y_k + a_{jK}, \quad (y_k \in [y_{k,j}^{\min}, y_{k,j}^{\max}]), \quad (2)$$

in the each terminal node, the processing error (1) can be estimated as:

$$\varepsilon_j(k, \alpha) = \min_a \left\| f_j^*(\bar{y}) - \bar{z}_j \right\|_{[y_{k,j}^{\min}, y_{k,j}^{\alpha}]} + \min_a \left\| f_j^{**}(\bar{y}) - \bar{z}_j \right\|_{[y_{k,j}^{\alpha}, y_{k,j}^{\max}]},$$

where \bar{a}^* , $f_j^*(\bar{y})$ – the regression coefficients and function (according to (2)) for the new left node;

\bar{a}^{**} , $f_j^{**}(\bar{y})$ – the regression coefficients and function (according to (2)) for the new right node;

α – the optimal threshold of partition;

k_j – the "best feature".

2) Calculation of the *regression coefficients* for each terminal vertex on the base of least squares method. We have to solve the system of linear algebraic equations by using all elements of the training dataset "fallen" into terminal vertex.

3) Checking the restrictions for computational complexity and estimation of processing quality on the base of test dataset.

3.2. Finishing of the construction process

To determine the stop parameters for algorithm construction process, we need to estimate the generalization capability of the local processing procedures.

There is a dataset: $\Omega = \{\bar{y}_j, z_j\}$, $j = \overline{1, T}$:

$$\Omega^T = \Omega^s \cup \Omega^t, \quad s+t=T, \quad \Omega^s \cap \Omega^t = \emptyset. \quad (3)$$

The quality of algorithm on a dataset:

$$v(a, \Omega) = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} I(\omega_i, a(\omega_i)),$$

where a – regression algorithm, A – set of algorithms, μ – method (algorithm) of teaching based on a dataset: $\mu(\Omega) = a$,

$$I = \begin{cases} 1, & |a(\omega_i) - \Phi(\omega_i)| \geq \delta(\omega_i) \\ 0, & |a(\omega_i) - \Phi(\omega_i)| < \delta(\omega_i) \end{cases}$$

Usually, for estimation of the generalization capability and limitations on the length of the training dataset the statistical theory is used [Vap74]. This theory is based on the analysis of functional of uniform deviation of the error rate in 2 samples:

$$P_\varepsilon^{st}(A) = P \left\{ \sup_{a \in A} (v(a, X^t) - v(a, X^s)) > \varepsilon \right\}.$$

But, the statistical theory characterized by heavy reliance of estimates and, moreover, this theory not take into account the possible shuffles of the elements of the training and testing samples. A more efficient way to estimate of the generalization capability is the usage of combinatorial theory [Vor04] based on a functional of a full sliding control, which is invariant to arbitrary permutations of samples:

$$Q^s(\mu, \Omega^T) = \frac{1}{N} \sum_{n=0}^{N-1} v(\mu(\Omega_n^s), \Omega_n^t), \quad (N = C_T^s), \quad (4)$$

and takes into account all three factors: the characteristics of samples distribution, restored dependence and the teaching method.

3.3. Construction process

Development of effective algorithm of the local image processing based on a hierarchical regression and such features as a local DWT of image requires simultaneous consideration of different performance indicators: the computational complexity of the procedure, its quality or processing error and generalization capability.

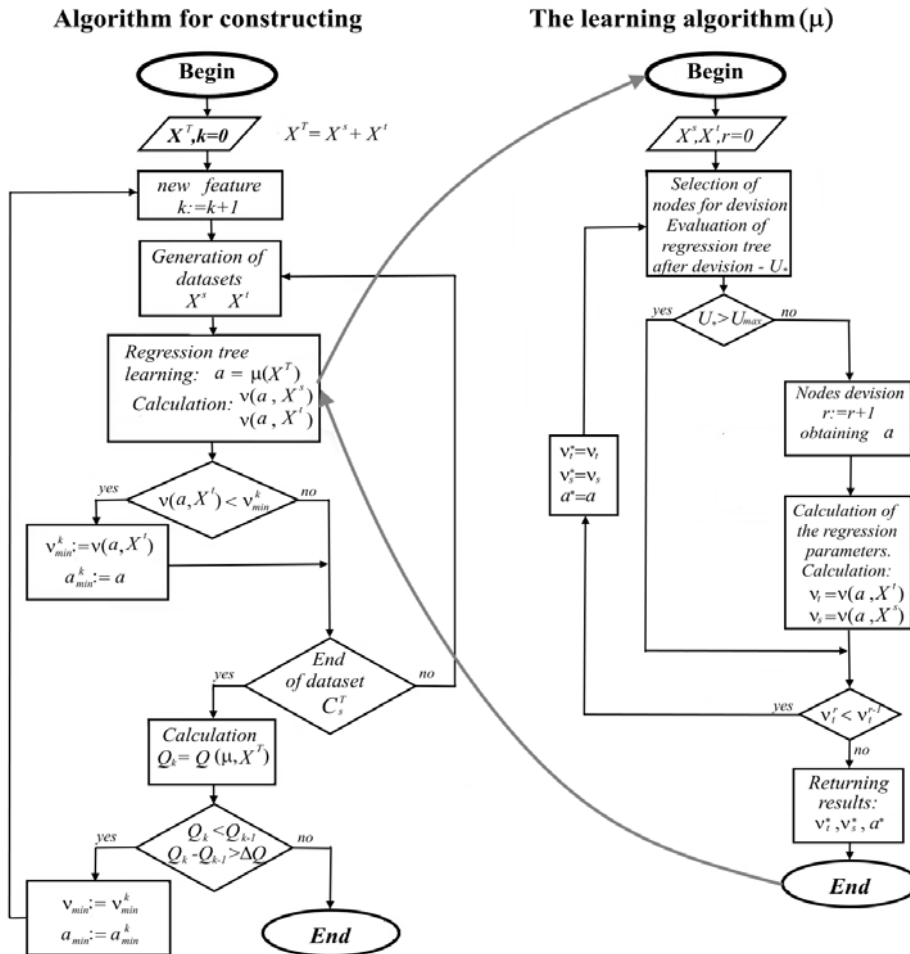


Figure 2. A diagram of algorithm for construction a computational procedure of local image processing.

The scheme of the algorithm of automatic construction of processing procedures is shown in Figure 2. The algorithm assumes the consistent accumulation of features as long as the functional of a full sliding control is decreasing (that means that a quality of processing is improving), and the computational complexity of the procedure remains in predetermined limits.

3.4. Experimental researches

As an experimental task, we consider the task of image filtration. The solution involves the use of local image processing procedures based on regression tree (RT) and artificial neural network

(NN). The comparison of the processing quality and computational complexity of these algorithms is presented in table 1. As can be seen from the table, the proposed method of hierarchical regression has the better accuracy with essentially smaller computational complexity than well known neural network method.

NN	ε	10.92	10.78	10.69	10.67	10.63	10.62	10.62	10.64
	U	54	99	189	279	369	459	549	621
RT	ε	11.28	11.01	10.89	10.81	10.72	10.62	10.57	10.61
	U	31	38	41	44	46	48	50	52

Table 1. The comparison results.

4. METHODOLOGY OF STOPPING OF TRAINING PROCESS.

4.1. Scheme of exhaustive search on datasets

Taking into account generalization capability of the local processing procedures based on a functional of a full sliding control [Vor04] we can estimate that the total number of all possible N decompositions of dataset is C_T^s . General scheme of construction procedure is shown in Figure 3.

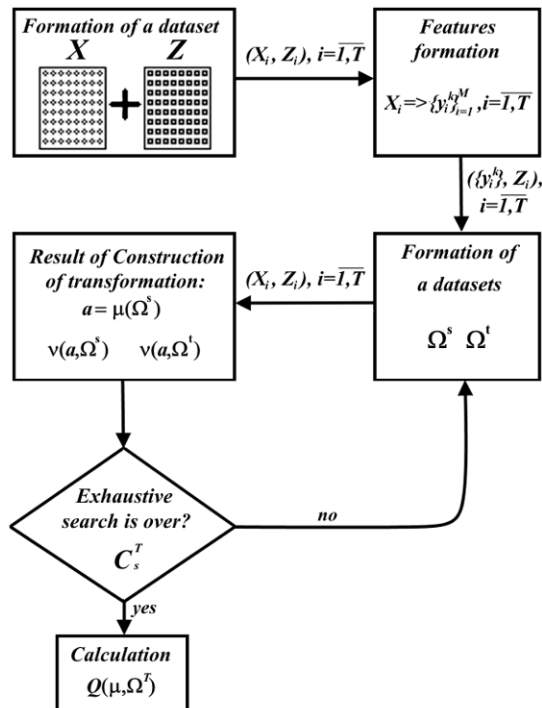


Figure 3. Schema of construction procedure with exhaustive search on datasets.

Is quite logical fact that in the case of images processing the construction of processing procedures which takes into account all combination of training and testing dataset is unrealizable because of the incredibly large busting on various combinations of datasets. Therefore, we have had to develop a method of determination rule for stop busting process on the base of a finite number of samples.

4.2. Scheme of exhaustive search on datasets

For sufficiently large sample volumes can be assumed that the error rate of the algorithm has a binomial distribution with t degrees of freedom (the length of the test dataset) and the probability of "success" = p (quality of the algorithm on a control set). In this way:

$$v(\mu(\Omega_n^s, \Omega_n^t) \sim Bin(t, p).$$

Function of the probability is specified as:

$$p_v(r) = C_t^r p^r (1-p)^{t-r}, \quad r = \overline{0, t}.$$

Then the distribution of the functional full cross-validation is evaluated by:

$$Q^{st}(\mu(\Omega), \Omega) = \frac{1}{N} \sum_{n=0}^{N-1} v(\mu(\Omega_n^s, \Omega_n^t) \sim Bin(N \cdot t, p).$$

Decision about continuing or stopping generation of different combinations training and control datasets and transition to the next subset of features can be taken based on the analysis of functional $Q_1^{st} \sim Bin(N_1 \cdot t, p_1)$, $Q_2^{st} \sim Bin(N_2 \cdot t, p_2)$ for the different subsets of features. We decide whether to recalculate the feature space, or to stop the process of building a processing procedure under the assumption of $p_2 < p_1$, with veracity γ (and correspondingly $p_1 < p_2$, with the veracity $(1-\gamma)$).

In such case the quality of the algorithm on dataset Ω can be estimated as:

$$v(\mu(\Omega), \Omega^T) \sim \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} I(\omega_i, \mu(\omega_i)),$$

$$\text{where } I(\omega_i, \mu(\omega_i)) = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}.$$

Moreover, if $n \gg 1$ (that is justified, because n is the number of objects and corresponds to the image size) and the λ is fixed, we obtain the Poisson distribution with parameter λ :

$$Bin(n, \lambda/n) \approx P(\lambda).$$

In this case to make a decision of stop generation process for various combinations of training and control datasets, and to transition to next features set we have to calculate confidence intervals for the expectation of a Poisson distribution for the functional full cross-validation on a datasets N_1 and N_2 in form:

$$\left[\lambda_1 - \frac{\tau_{1-\alpha/2} \sqrt{\lambda_1}}{\sqrt{N_1}}, \lambda_1 + \frac{\tau_{1-\alpha/2} \sqrt{\lambda_1}}{\sqrt{N_1}} \right] \left[\lambda_2 - \frac{\tau_{1-\alpha/2} \sqrt{\lambda_2}}{\sqrt{N_2}}, \lambda_2 + \frac{\tau_{1-\alpha/2} \sqrt{\lambda_2}}{\sqrt{N_2}} \right],$$

where $\tau_{1-\alpha/2}$ – quantile of distribution $N_{0,1}$ for level $1-\alpha/2$ ($\alpha = 1-\gamma$).

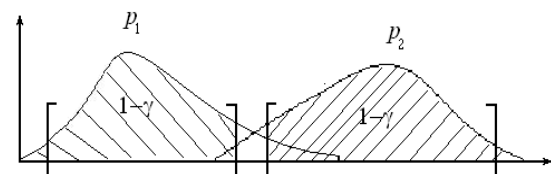


Figure 4. Calculation of confidence intervals.

The decision of stopping generation of different combinations training and control datasets and about the switching to the next subset of features is taken at

a moment when a separation of calculated confidence intervals on adjacent steps is achieved.

4.3. Illustration of the process of processing algorithm construction

Figure 5 shows an example of training of a regression tree, for different sets of features ($K=1,2,3,\dots,12$, with a gradual increase). The graphs show the noise reduction (ε^2/D_V) with the increasing of a regression tree depth (H_{av}). Figure 6 presents a statistics of process of regression tree construction on a various combinations of training and control datasets (group of points of each colors correspond to the optimum value of quality for a given set of features $K=1,2,3,\dots,12$, in the case of exhaustive search a some number of partitioning of a dataset Ω on training and control part (Ω_n^s, Ω_n^t), $n = 1, 2, \dots, N$).

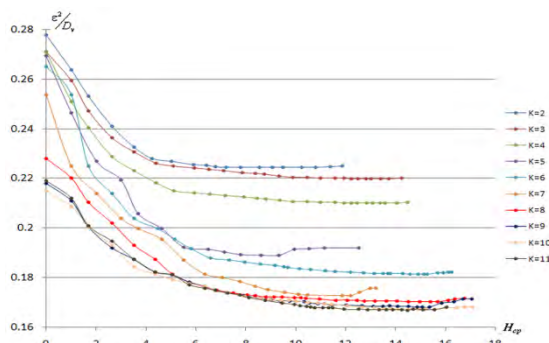


Figure 5. Process of training of processing procedure at various features space.

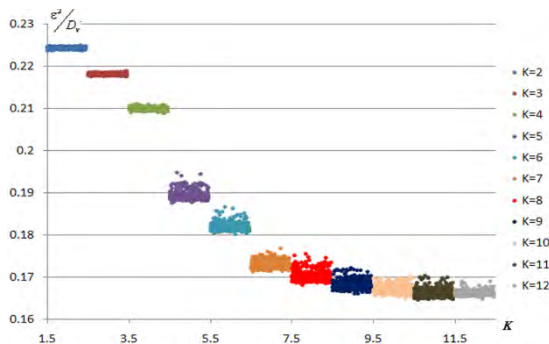


Figure 6. Statistics of quality of procedures for different combinations of training and control datasets.

Figure 7 shows a graph of the construction process of local image processing procedures, with confidence intervals, for the optimal values of quality, and Figure 8 – calculation of the required number of combinations of training/testing datasets for the making a decision of switching to the next set of features (the number of combinations required for the separation of the confidence intervals on the adjacent steps).

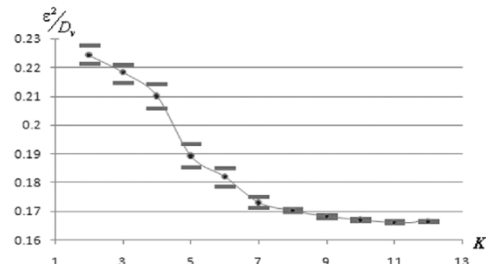


Figure 7. Construction of the processing procedure with confidence intervals.

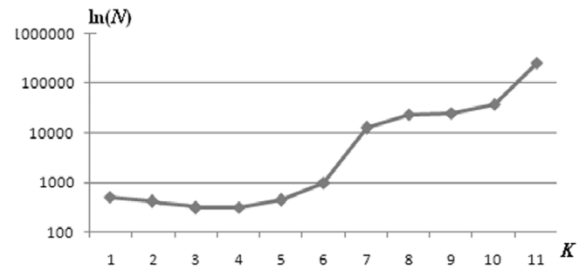


Figure 8. Calculating of combinations number.

5. CONCLUSIONS AND RESULTS

The paper presents an *efficient technology* that allows to realize automatic construction of computational procedure of local processing of digital signals/images. In accordance with the creation processes the constructed computational procedure has a specified complexity and the highest quality and the generalizing ability. The proposed method of estimation of the required number of algorithm training iterations and, as a consequence, the stopping rule of the formation different combinations of training and testing datasets based on their particular number allows to use the full functionality of combinatorial theory and a functional of full cross-validation control during the constructing (training) processing procedures, which are tuning on the bases of a training dataset. And as a result, possible to prevent problems of retraining/poorly trained processing algorithms and, at the same time, construct the local processing procedure with predetermined computational complexity and veracity, and with the best quality (for an existing training dataset).

6. ACKNOWLEDGEMENTS

This work was financially supported by the Russian Scientific Foundation (RSF), grant no. 14-31-00014 “Establishment of a Laboratory of Advanced Technology for Earth Remote Sensing”.

7. REFERENCES

- [Soi09] Methods of computer image processing. Part II: Methods and algorithms / M.V. Gashnikov, N.I. Glumov, N.Yu. Ilyasova, V.V. Myasnikov, S.B. Popov, V.V. Sergeev, V.A. Soifer,

- A.G. Hramov, A.V. Chernov, V.M. Chernov, M.A. Chicheva, V.A. Fursov. – Ed. by V.A. Soifer. – Moscow: “Fizmatlit” Publisher, 2009. – 784 p.
- [Woo05] Woods, R. Digital Image Processing / R. Woods, R.Gonsales / - M: Technosphere, 2005. - 1072 p. (in russian)
- [Pra07] Pratt, W. Digital image processing. Wiley, 4ed, 2007.
- [Hai06] Haikin, S. Neural Networks: A Comprehensive Foundation / M.: «Vilyams», 2006. 1104 p.
- [Bre84] Breiman L. Classification and regression trees / Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone.// Monterey, Calif., U.S.A.: Wadsworth, Inc. – 1984.
- [Kop12] Kopenkov V.N., Myasnikov V.V. An algorithm for automatic construction of a local non-linear processing procedures images based on hierarchical regression. Computer optics, 36(2):257-266, 2012. (in russian).
- [Vap74] Vapnik, V.N. The theory of pattern recognition / VN Vapnik, A. Chervonenkis / - Moscow: Nauka, 1974. (in russian)
- [Vor04] Vorontsov K. A combinatorial approach to assessing the quality of training algorithms // Mathematical problems of cybernetics / Ed O.B. Lupanov. Moskow. Fizmatlid. 2004. Vol. 13. p. 5–36. (in russian)
- [Kop08] Kopenkov V. Efficient algorithms of local discrete wavelet transform with HAAR-like bases. Pattern Recognition and Image Analysis. Vol 18 No 4 2008 pp. 654-661.
- [Kop14] Kopenkov V. On halting the process of hierarchical regression construction when implementing computational procedures for local image processing. Pattern Recognition and Image Analysis. Vol 24 No 4 2014 pp. 506–510