# PATTERN RECOGNITION AND AUTOMATED CLASSIFICATION OF MULTICOMPONENT VIDEO IMAGES UNDER CONDITION OF STATISTICAL INDETERMINACY

**K.T. Protasov**

*Institute of Atmospheric Optics,*
*Siberian Branch of the Russian Academy of Sciences, Tomsk*
*Received December 6, 1994*

*Within the context of a problem of monitoring of the Earth's underlying surface and cloudiness, new estimating Bayes decision rules for pattern recognition and automated classification were synthesized to extract complete information from a set of satellite video images as well as cartographic information and results of contact measurements under conditions of a priori indeterminacy. The rules take into account high dimensionality of miscellaneous observations, meager tutorial sequences, and degeneracy of support functions of approximating distributions.*

In order to assess the state of the Earth's underlying surface and cloudiness, one should process a large quantity of satellite multidimensional data recorded in tens of spectral channels of visible (V), infrared (IR), and superhigh–frequency (SHF) ranges of electromagnetic radiation taking into account results of subsatellite contact measurements and landscape subject–oriented mapping. The necessity of taking into account the temporal characteristics and spatial texture of video images for solving the problems of exploration of natural resources and ecological monitoring increases the dimensionality of observations to be analyzed. The development of techniques for fusion of miscellaneous images and the advent of satellite means of periodical observation of the Earth's surface and cloudiness open the prospects for joint analysis of miscellaneous data obtained at different times. Traditional approach to image analysis is oriented at processing of such individual components as a set of V, IR, or radar (R) images with subsequent aggregation of component solutions. This leads to irretrievable loss of information on inter–component connections.

The problem arises of the complete extraction of the information from multicomponent miscellaneous data of essentially high dimensionality.

Among the algorithms of preliminary video image processing, the main role is assigned to the pattern recognition algorithms with "instruction" on test image fragments and to the algorithms of automated classification when tutorial samples are lacking. These algorithms make it possible to analyze the geometry of multidimensional sample spaces and to solve effectively the problem of aggregation, objective segmentation, and identification of the data.

Taking into account the stochastic character of recorded fields, to create the algorithms for the pattern recognition and automated classification, it is natural to apply the statistical theory of hypothesis testing and parameter estimation based on the conditional probability distribution of recognizable situations that, as a rule, are unknown.

In the last few years, the theory is successfully developed of nonparametric estimation of the unknown distribution from data samples, but the questions of the choice of a kernel for these estimates and, what is more essential, of the smoothing parameters are still open questions.

Let us define the structure of the observed miscellaneous data and introduce the term of the observation element, namely, the aggregate of data. Let us suppose that all the recorded image components of V, IR, and SHF ranges, accompanying cartographic data, and the data of contact measurements have been digitized, scaled, and normalized so that a choice of a fragment of $m_x \times m_y$ pixels of the brightness characteristics, for example, of visible spectral range as an observational element to consider the texture characteristics allows us to generate a series of corresponding fragments in all the spectral channels of V, IR, and SHF ranges, and to complete it by the fragments of cartographic data and the vector of contact measurements. In particular, the hexagonal fragment of the spatial information medium connected with cartographic basis can be selected depending on the problem to be solved.

Depending on a specific problem, so–formatted aggregate of observational data can be expressed in the form of the following structure with multi–indices:

$$\mathbf{z(u)}=\left\{\begin{array}{ll}\{x^{ijk}\}m_i\times m_j\times m_k, & i=1,\ldots,m_i,\ j=1,\ \ldots,m_j,\ k=1,\ldots,\ m_k,\\ \{y^{\mu\nu}\}\ m_\mu\times m_\nu, & \mu=1,\ \ldots,\ m_\mu,\ \nu=1,\ \ldots,\ m_\nu,\\ \{v^l\}m_l, & l=1,\ \ldots,\ m_l,\end{array}\right\},$$

where $\{x^{ijk}\}$ are the digitized brightness components of V, IR, and SHF ranges with the values of $m_i \times m_j$ square of pixels that constitute the reference fragment of video images; $m_k$ is the number of individual images; $\{y^{\mu\nu}\}$ are the digitized cartographic data with the values of $m_\mu \times m_\nu$ square of pixels; and, $\{v^l\}$ is the $m_l$–dimensional vector of subsatellite observations.

If we designate the generalized dimensionality of the aggregate of data by $n$, where $n = m_i \times m_j \times m_k + m_\mu \times m_\nu + m_l$, then $\mathbf{z(u)}$, where $\mathbf{u}$ is a series of multi–indices that may sometimes be omitted, is the element of the $n$-dimensional Euclidean space $\mathbf{z(u)} \in E^n$. Although, from theoretical standpoint, the space of multidimensional matrix structures of data is isomorphic with the space of vectors whose components are composed of all the components of the matrix data, nevertheless, we will not vector the aggregates $\mathbf{z(u)}$ because vector representation destroys the natural data structure. In this connection, it is necessary to determine

0235-6880/95/06 430-05 $02.00 

the scalar (inner) and direct (outer) products of the elements $\mathbf{z}(\mathbf{u})$ and $\mathbf{z}'(\mathbf{u}) \in E^n$. For compactness, vector notation of operations and corresponding terms explained in Appendix will be used.

Let us consider the problem of formulation of the decision rule of the pattern recognition in the statistical statement, when tutorial sequences of aggregates obtained from test fragments of video images and supplemented by accompanying data of contact measurements have been given.

Let the probability measures with *a priori* distribution of situations $P(\lambda)$ and conditional probability density functions $f_\lambda(\mathbf{z}(\,))$ of the aggregates of observational data $\mathbf{Z}(\,) \in E^n$ be in the $n$-dimensional Euclidean space $E^n$ of the aggregates of data $\mathbf{z}(\,) \in E^n$ and the image space $\Lambda = \{1, \ldots, L\}$ of $L$ patterns (classes), where $n$ is the generalized dimensionality of data. Let us determine the simple matrix of losses due to accepted decisions $(1 - \delta_{\lambda\mu})$, where $\delta_{\lambda\mu}$ is the Kronecker symbol.

As is known,[1] the Bayes decision rule optimal in the sense of mean loss minimum has the following form:

$$r(\mathbf{z}(\,)) = \arg\max_{\lambda \in \Lambda} P(\lambda)\, f_\lambda(\mathbf{z}(\,)), \qquad r, \lambda \in \Lambda, \tag{1}$$

where the decision $r(\,)$ also belongs to the class space $\Lambda$. The conditional density functions $f_\lambda(\,)$ are unknown for real problems, but there are the tutorial sequences of data samples classified by the "tutor" $\mathbf{z}^\lambda_1(\,), \ldots, \mathbf{z}^\lambda_{N_\lambda}(\,)$, where $N_\lambda$ is the size of a sample belonging to the class $\lambda \in \Lambda$. To retrieve the unknown distributions in $E^n$ it is natural to use their nonparametric estimates, for example, in the metric with the Gauss kernel[2]:

$$\hat{f}_\lambda(\mathbf{z}(\,)) = \frac{N_\lambda^{-1}\, h_\lambda^{-n}}{(\sqrt{2\pi})^n\, |\hat{G}_\lambda|^{1/2}} \times$$

$$\times \sum_{j=1}^{N_\lambda} \exp\left\{-\frac{1}{2\, h_\lambda^2} (\mathbf{z}(\,) - \mathbf{z}^\lambda_j(\,))^{\mathrm{T}}\, \hat{G}_\lambda^{-1}\, (\mathbf{z}(\,) - \mathbf{z}^\lambda_j(\,))\right\}, \tag{2}$$

where $\hat{G}_\lambda$ is the correlation matrix estimated from the tutorial sample of the class, T is the transposition symbol, $h_\lambda$ is the smoothing parameter, $N_\lambda$ is the size of tutorial sample, $\lambda \in \Lambda$. Taking into account the degeneracy of the inverse correlation matrix $\hat{G}_\lambda^{-1}$ due to insufficient statistic of tutorial data samples for $n \geq N_\lambda$, let us consider some questions of correct calculation of $\hat{f}_\lambda(\mathbf{z}(\,))$ using the results from Appendix.

To this end, we represent the quadratic form of the Gauss kernel in Eq. (2) as follows [see Eq. (A6)]:

$$Q_j = (\mathbf{z}(\,) - \mathbf{z}_j(\,))^{\mathrm{T}}\, \hat{G}^{-1}\, (\mathbf{z}(\,) - \mathbf{z}_j(\,)) = (\mathring{\mathbf{z}}(\mathbf{u}) - \mathring{\mathbf{z}}_j(\mathbf{u}))^{\mathrm{T}} \times$$

$$\times \sum_{i=1}^{k} \frac{1}{\sigma_i^2}\, \boldsymbol{\Phi}_i(\mathbf{u})\, \boldsymbol{\Phi}_i^{\mathrm{T}}(\mathbf{v})\, (\mathring{\mathbf{z}}(\mathbf{v}) - \mathring{\mathbf{z}}_j(\mathbf{v})) + (\mathring{\mathbf{z}}(\mathbf{u}) - \mathring{\mathbf{z}}_j(\mathbf{u}))^{\mathrm{T}} \times$$

$$\times \sum_{i=k+1}^{n} \frac{1}{\sigma^2}\, \boldsymbol{\Phi}_i(\mathbf{u})\, \boldsymbol{\Phi}_i^{\mathrm{T}}(\mathbf{v})\, (\mathring{\mathbf{z}}(\mathbf{v}) - \mathring{\mathbf{z}}_j(\mathbf{v})),$$

where the aggregates are centered around the estimate of mathematical expectation of corresponding class, $\mathring{\mathbf{z}} = \mathbf{z}(\,) - \hat{\boldsymbol{\mu}}(\,)$, $\sigma_i^2 = \lambda_i$, $\{\boldsymbol{\Phi}_i(\,)\}_1^k$ is the dominant part [see Eq. (A6)] and $\{\boldsymbol{\Phi}_i(\,)\}_{k+1}^n$ is the additive part of the Karhunen–Loeve basis, $\sigma$ is the regularization parameter invariable in all additive basis due to the lack of information for its concrete definition.

Let us transform the quadratic form $Q_j$ taking into account the following remarks. First, since the dominant part of the basis $\{\boldsymbol{\Phi}_i(\,)\}_1^k$ was sampled from the class, we can represent the centered values of the tutorial sample in the basis of their proper class to a high accuracy. In this connection, the coefficients of representation of these sampled values in the additive basis $\{\boldsymbol{\Phi}_i(\,)\}_{k+1}^n$ are close to zero, and we can ignore them. We then obtain

$$Q_j = \sum_{i=1}^{k} \frac{(y^i - x_j^i)^2}{\sigma_i^2} + \sum_{i=k+1}^{n} \frac{(y^i)^2}{\sigma^2},$$

where $y^i$ and $x_j^i$ are the coefficients of representation of the centered observations $\mathring{\mathbf{z}}(\,)$ and $\mathring{\mathbf{z}}_j(\,)$ in the bases $\{\boldsymbol{\Phi}_i(\,)\}_1^k$ and $\{\boldsymbol{\Phi}_i(\,)\}_{k+1}^n$ [see Eq. (A3)]. It is easy to see that $\sum_{i=k+1}^{n} (y^i)^2$ is the error in approximation of the observed realization $\mathbf{z}(\,)$, preliminary centered about the mathematical expectation of the class $\hat{\boldsymbol{\mu}}(\,)$ in the basis of the same class, i.e.,

$$\sum_{i=k+1}^{n} (y^i)^2 = \langle \mathring{\mathbf{z}}(\mathbf{u}) - \sum_{i=1}^{k} y^i\, \boldsymbol{\Phi}_i(\mathbf{u}) \rangle^2 = \mathcal{E}_k^2\, (\mathring{\mathbf{z}}(\,)).$$

Thus, the modified estimate of the unknown density function has the form:

$$\tilde{f}_\lambda(\mathbf{z}(\,)) = \frac{1}{(\sqrt{2\pi})^{n-k_\lambda}} \frac{1}{\sigma_\lambda^{n-k_\lambda}} \exp\left\{-\frac{1}{2\, \sigma_\lambda^2}\, \mathcal{E}_{k_\lambda}^2\, (\mathring{\mathbf{z}}(\,))\right\} \times$$

$$\times \frac{N_\lambda^{-1}}{(\sqrt{2\pi})^{k_\lambda}} \frac{h_\lambda^{-k_\lambda}}{\left(\prod_{i=1}^{k_\lambda} \sigma_{\lambda i}\right)} \sum_{j=1}^{N_\lambda} \exp\left\{-\frac{1}{2\, h_\lambda^2} \sum_{i=1}^{k_\lambda} \frac{(y^i - x_j^i)^2}{\sigma_{\lambda i}^2}\right\} \tag{3}$$

and the following geometric interpretation: in the data sample space we have a standard nonparametric density estimate in the space of the spectral parameters of the Karhunen–Loeve basis (the second component in Eq. (3)). Due to deficient information, the first component in Eq. (3) is the Gauss density function describing the distribution of the error in representing the arbitrary observation in the basis of the class. Let us characterize the quality of decision rule (1) with modified estimates of the density functions given by Eq. (3) by the empirical risk[7], namely

$$\hat{R} = \sum_{\lambda \in \Lambda} \frac{1}{N_\lambda} \sum_{j=1}^{N_\lambda} P(\lambda)\, I\, \{\lambda = \arg\max_{\mu \in \Lambda} P(\mu)\, \tilde{f}_\mu(\mathbf{z}_j^\lambda(\,))\}, \tag{4}$$

where $I\{\text{truth}\} = 0$, and $I\{\text{falsehood}\} = 1$ is the indicator function of false decisions. The empirical risk is calculated by the slide control technique to save the sample values. Namely, when calculating $\tilde{f}_{\mu=\lambda}(\mathbf{z}_j^\lambda(\,))$ at the point $\mathbf{z}_j^\lambda(\,)$ for $\lambda = \mu$ in Eq. (4), this point is excluded from the data sample $\mathbf{z}_1^\lambda(\,), \ldots, \mathbf{z}_{j-1}^\lambda, \mathbf{z}_{j+1}^\lambda(\,), \ldots, \mathbf{z}_{N_k}^\lambda(\,)$ used to estimate the function $\tilde{f}_{\mu=\lambda}(\,)$ (see Ref. 3).

It should be noted that each density function estimate $\tilde{f}_\lambda(\ )$ has two subdefinite parameters $h_\lambda$ and $\sigma_\lambda$, $\lambda \in \Lambda$. It is natural to select such parameters $h_\lambda$ and $\sigma_\lambda$, $\lambda \in \Lambda$ that minimize the risk function $\hat{R}$ given by Eq. (4). Taking into account that $\hat{R}$ has many extrema and is not differentiable, to solve the minimization problem we have modified the search methods of optimization[5] that combine random search with local steepest descent.

In this case, it i necessary to know the range of variation of the adjustable parameters. We can estimate the approximate extreme values of the parameters from the condition of maximum of the empirical likelihood functional,[6] then

$$\left[\sum_{i=1}^{N_\lambda} \min_{\{i\}} \sum_{j|l=1}^{k_\lambda} \frac{(x_i^l - x_j^l)^2}{\sigma_{\lambda l}^2 N_\lambda k_\lambda}\right]^{1/2} < h_\lambda < \left[\sum_{i=1}^{N_\lambda} \max_{\{i\}} \sum_{l=1}^{k_\lambda} \frac{(x_i^l - x_j^l)^2}{\sigma_{\lambda l}^2 N_\lambda k_\lambda}\right]^{1/2}, \tag{5}$$

$$\sigma_\lambda \cong \left[\frac{\sum_{j=1}^{M_\lambda} \mathcal{E}_{k_\lambda}^2(\mathring{\mathbf{z}}_j(\ ))}{(n - k_\lambda)\, M_\lambda}\right]^{1/2},$$

where $\mathring{\mathbf{z}}_1^\lambda(\ ), \ldots, \mathring{\mathbf{z}}_{M_\lambda}^\lambda(\ )$ is the tutorial sample of all classes of the size $M_\lambda = \sum_{i=1, i\neq\lambda}^{\Lambda} N_i$, when the class $\lambda$ is removed from it. The sample is centered around mathematical expectation of the class $\lambda$, and the error $\varepsilon_{k_\lambda}^2(\ )$ is calculated in the basis of the class $\lambda \in \Lambda$.

Once the parameters $h_\lambda$ and $\sigma_\lambda$, $\lambda \in \Lambda$, minimizing the empirical risk, have been determined, decision rule (1) can be used for an analysis of aggregates of data that are not included in instruction of the algorithm.

Let us now suppose that tutorial samples are lacking, but there is a mixed sample, and the problem is to identify a few compact groups of data samples, called clusters or taxons, studying the geometry of data arrangement in the mixed sample. The classes of compactness or more exactly, sample values composing these classes, can serve as tutorial samples to formulate estimating Bayes decision rules (1).

Let us take a statistical model describing the situation. Suppose we have a nonclassified sample of the aggregates of observations $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$, where $N$ is the sample size, $\mathbf{z}(\ ) \in E^n$, at our disposal.

We suppose that the random aggregate $\mathbf{Z}(\ )$ has the function of the probability density of the form:

$$f(\mathbf{z}(\ )) = \sum_{\lambda\in\{1,\ldots,L\}} P(\lambda)\, f_\lambda(\mathbf{z}(\ )), \tag{6}$$

where $L$ is the number of clusters, $f_\lambda(\mathbf{z}(\ ))$ is the conditional unimodular density function of the cluster $\lambda$, $P(\lambda)$ is the weight of the density function $f_\lambda(\ )$ in the mixture that means the *a priori* probability of occurrence of the cluster $\lambda$, $\sum_{\lambda\in\{1,\ldots,L\}} P(\lambda) = 1$, and the quantities included into Eq. (6) are unknown.

The problem is to reconstruct the components $\{L, P(\lambda), f_\lambda(\ ), \lambda \in \{1, \ldots, L\}\}$ of mixture (6) from the available nonclassified sample $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$ of size $N$ of the observed aggregates $\mathbf{z}(\ )$. It should be noted that the problem of reconstruction of the components of mixture (6) is solvable only when it is identifiable.[7] It is difficult to check this condition in practice, and from the geometric standpoint it means that $f(\mathbf{z}(\ ))$ should have well pronounced local modes engendered by cluster–induced subsamples of mixed sample. In addition, to simplify the problem we suppose that the size of these cluster–induced subsamples are approximately proportional to $P(\lambda)$, and

$$P(\lambda) \cong N_\lambda / N\ ,\ \lambda \in \{1, \ldots, L\}\ .$$

Solving the problem of reconstruction of the mixture in the correct statement is connected with a search for the maximum of the likelihood functional by varying the unknown parameters taking into account the chosen parameterization $f_\lambda(\ )$ and is too cumbersome for calculations. So heuristic approaches to approximate solution of this problem on the basis of analysis of the sample aggregate geometry in the space of the mixed sample are justified.

The hierarchical procedure of cluster identification is constructed as follows: the "distance" function is defined in the space of observations, and the distances between all possible parameters of the data sample $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$ are analyzed. At the first step, the pairs of elements that are closest to each other in the sense of selected distance[8] are clustered. The result of the first step is identification of the cluster centers of the aggregates. At the second and subsequent steps, all distances between clusters identified at the first step are analyzed, with the distance between clusters being defined as a distance between one cluster and each point of another cluster, the nearest point of the second cluster is selected, the nearest clusters are pooled, then clusters are pooled again, and their number decreases.

Clusters are pooled until their necessary number or the number close to expected one is obtained or all the points of the mixed sample are pooled into one taxon.

Let $\mathbf{z}_1^\lambda(\ ), \ldots, \mathbf{z}_{M_\lambda}^\lambda(\ )$ be the subsample of the size $M_\lambda$ from the mixed sample $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$ identified as the cluster $\lambda$ at some iteration step. We define the probability that the examined observation $\mathbf{z}(\ )$ belongs to the aggregate $\mathbf{z}_1^\lambda(\ ), \ldots, \mathbf{z}_{M_\lambda}^\lambda(\ )$ identified as a cluster of observations as a nonparametric estimate of the density function at the point $\mathbf{z}(\ )$

$$\rho(\mathbf{z}(\ );\, \mathbf{z}_1^\lambda(\ ), \ldots, \mathbf{z}_{M_\lambda}^\lambda(\ )) = \frac{1}{N} \sum_{j=1}^{M_\lambda} \frac{|\hat{G}|^{-1/2}\, h_\lambda^{-n}}{(\sqrt{2\pi})^n} \times$$

$$\times \exp\left\{-\frac{1}{2\, h_\lambda^2} (\mathbf{z}(\ ) - \mathbf{z}_j^\lambda(\ ))^{\mathrm{T}}\, \hat{G}^{-1}\, (\mathbf{z}(\ ) - \mathbf{z}_j^\lambda(\ ))\right\}, \tag{7}$$

where $h_\lambda$ is the smoothing parameter varying during the iteration process from $h_{\min}$ to $h_{\max}$; $\hat{G}$ and $\hat{G}^{-1}$ are the estimates of correlation and inverse correlation matrices calculated from the sample $\mathring{\mathbf{z}}_1^\lambda(\ ), \ldots, \mathring{\mathbf{z}}_{M_\lambda}^\lambda(\ )$ of the size $M_\lambda$ taking into account the stability of calculation. We note that at the first steps of the iteration procedure when the size $M_\lambda$ of the data sample of the cluster is small, to estimate $\hat{G}$ and $\hat{G}^{-1}$, we should use the entire mixed sample. Once all distances $\rho_\lambda(\mathbf{z}(\ ))$ have been calculated, where $\lambda \in \{1, \ldots\}$, the element $\mathbf{z}(\ )$ is added to the set $\{\mathbf{z}_j^\lambda\}_1^{M_\lambda}$ for which the distance $\rho_\lambda(\ )$ is maximum, in other words, by the Bayes rule given by Eq. (1) with estimates $\tilde{P}(\mathrm{l})$ and $\tilde{f}_\lambda(\mathbf{z}(\ ))$ given by Eq. (7).

To measure distances in the sense of Eq. (7), the smoothing parameter is varied within some limits or is set as follows[6]:

$$h = \left[ \frac{\sum_{i=1}^{N} \sum_{j \neq i}^{N} (\mathbf{z}_i(\ ) - \mathbf{z}_j(\ ))^{\mathrm{T}} \hat{G}^{-1} (\mathbf{z}_i(\ ) - \mathbf{z}_i(\ ))}{N(N-1)\, n} \right]^{1/2}. \quad (8)$$

Thus, if the mixture (6) is identifiable and the local extrema, i.e., the modes of the cumulative density function, are quite "well pronounced", the clusters of the data sample identified during the above–described iteration process reconstruct the unimodular components of the mixture observed.[9]

## APPENDIX

Using the agreement on the vector–functional form of writing the aggregates $\mathbf{z}(\ ) = \mathbf{z}(\mathbf{u})$ and $\mathbf{w}(\ ) = \mathbf{w}(\mathbf{u})$, we define their inner or scalar product as follows:

$$\mathbf{z}^{\mathrm{T}}(\ )\, \mathbf{w}(\ ) = (\mathbf{z}(\ ),\, \mathbf{w}(\ )) = \sum_{\{\mathbf{u}\}} \mathbf{z}(\mathbf{u})\, \mathbf{w}(\mathbf{u}),$$

where the product are formed by the multipliers that are the corresponding components of the aggregates $\mathbf{z}(\mathbf{u})$ and $\mathbf{w}(\mathbf{u})$, and the summation is made over all pairs of identical multi–indices.

We define the outer or direct product of the aggregates $\mathbf{z}(\ )$ and $\mathbf{w}(\ )$ as follows:

$$\mathbf{z}(\mathbf{u})\, \mathbf{w}^{\mathrm{T}}(\mathbf{v}) = \mathbf{z}(\mathbf{u}) \otimes \mathbf{w}(\mathbf{v}),$$

then the new structure object with dimensionality $n \times n$ appears. This structure consists of the elements formed by products of all possible component pairs from $\mathbf{z}(\mathbf{u})$ and $\mathbf{w}(\mathbf{u})$, and each element of this structure is labeled by complex index consisting of the components of $\mathbf{u}$ and $\mathbf{v}$, respectively.

Now we take advantage of the technique for constructing the Karhunen–Loeve basis from the data samples[3] when observations are the aggregates of miscellaneous data $\mathbf{z}(\ )$. We suppose that the data sample $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$, where $\mathbf{z} \in E^n$, of size $N$ being the result of $N$ observations of the random aggregate $\mathbf{Z}(\ )$, is at our disposal. Let us suppose that the sample is centered around the selected mathematical expectation.

Let us represent the observed aggregate as follows:

$$\mathbf{Z}(\ ) \cong \sum_{i=1}^{k} X^i\, \mathbf{\Phi}_i(\ ), \quad (A.1)$$

where the random coefficients $\{X^i\}_1^k$ and the nonrandom basis of orthonormal functions–aggregates $\{\mathbf{\Phi}_i(\ )\}_1^k$ are determined from the conditions of minimization of the average square criterion of the approximation quality of the random aggregate $\mathbf{Z}(\ )$ in this basis

$$\mathcal{E}_k^2 = \mathrm{M} \parallel \mathbf{Z}(\ ) - \sum_{i=1}^{k} X^i \mathbf{\Phi}_i(\ ) \parallel^2 = \min(\{X^i\}_1^k, \{\mathbf{\Phi}_i(\ )_1^k\}), \quad (A.2)$$

where M is the symbol of the mathematical expectation operator, $\| \cdot \|$ is the Euclidean norm in the aggregate space $E^n$.

Coefficients $\{X^i\}_1^k$ of Eq. (1) minimizing (A.2) on the fixed basis $\{\mathbf{\Phi}_i(\ )\}_1^k$ have the following form:

$$X^i = (\mathbf{Z}(\ ),\, \mathbf{\Phi}_i(\ )) = \sum_{\{\mathbf{u}\}} \mathbf{Z}(\mathbf{u})\, \mathbf{\Phi}_i(\mathbf{u}), \quad i = 1, \ldots, k, \quad (A.3)$$

where $(,)$ is the symbol of the above introduced scalar product. Solving the above–indicated variational problem given by Eq. (A.2) for the conditional (in the sense of constraints of orthonormality of the basis elements $\{\mathbf{\Phi}_i(\ )\}_1^k$ imposed upon functional (A.2) via the Lagrange multipliers) extremum leads to a homogeneous Fredholm integral equation of the second kind

$$(\mathrm{M}[\mathbf{Z}(\mathbf{u})\, \mathbf{Z}^{\mathrm{T}}(\mathbf{v})],\, \mathbf{\Phi}(\mathbf{v})) = \lambda\, \mathbf{\Phi}(\mathbf{u}), \quad (A.4)$$

where $\lambda$ is the Langrange multiplier, and the indices of basis elements $\mathbf{\Phi}(\ )$ and $\lambda$ are omitted because of the equivalence of all equations.

If we use the estimate of the correlation function from a sample of the centered aggregates $\mathbf{z}_1(\ ), \ldots, \mathbf{z}_N(\ )$

$$\mathrm{M}[\mathbf{Z}(\mathbf{u})\, \mathbf{Z}^{\mathrm{T}}(\mathbf{v})] \cong \frac{1}{N} \sum_{j=1}^{N} \mathbf{z}_j(\mathbf{u}) \otimes \mathbf{z}(\mathbf{v}), \quad (A.5)$$

then the problem of finding the basis functions from Eq. (A.4) becomes essentially simpler and reduces to solving the general problem of the eigenvalues of the positively determined Gram matrix[3] of order $N$. As a result, we can determine the Karhunen–Loeve basis $\{\mathbf{\Phi}_i(\ )\}_1^k$ and the spectrum of the eigenvalues $\{\lambda_i\}_1^k$, $k \le N$, where $k$ is the serial number of the last "stable" eigenvalue of the spectral components placed in decreasing order $\lambda_1 \ge \ldots \ge \lambda_k \ge \ldots \ge \lambda_N$.

Using only stable eigenvalues and Mercer theorem,[4] we can estimate the inverse correlation function as follows:

$$\{\mathrm{M}[\mathbf{Z}(\mathbf{u})\, \mathbf{Z}^{\mathrm{T}}(\mathbf{v})]\}^{-1} \tilde{=} \sum_{i=1}^{k} \frac{1}{\lambda_i}\, \mathbf{\Phi}_i(\mathbf{u})\, \mathbf{\Phi}_i^{\mathrm{T}}(\mathbf{v}), \quad (A.6)$$

that is used in the Gauss metric for constructing the nonparametric estimates of the density function.

## REFERENCES


1. V.G. Repin and G.P. Tartakovskii, *Statistical Analysis under Conditions of a priory Indeterminacy and Adaptation of Informational Systems* (Sov. Radio, Moscow, 1977), 432 pp.
2. K. Fukunaga, *Introduction into the Statistical Theory of Pattern Recognition* [Russian Translation] (Nauka, Moscow, 1979), 368 pp.
3. K.T. Protasov, Atmos. Opt. **3**, No. 1, 42–46 (1990).
4. G. Van Tris, *Theory of Detection, Estimates, and Modulation* [Russian Translation] (Sov. Radio, Moscow, 1972), Vol. 1, 744 pp.
5. Ya.Z. Tsypkin, *Adaptation and Instruction in Automated Systems* (Nauka, Moscow,1968), 400 pp.
6. N.V. Ivanova and K.T. Protasov, in: *Mathematical Statistics and its Applications,* No. 8 (State University Publishing House, Tomsk, 1982), pp. 50–65.
7. S.A. Aivazyan, K.M. Bukhshtaber, I.S. Enyukov, and L.D. Meshalkin, *Applied Statistics: Classification and Dimension Decrease* (Financy i Statistika, Moscow, 1989), 607 pp.
8. E.A. Patric, *Principles of the Pattern Recognition Theory* (Sov. Radio, Moscow, 1980), 408 pp.
9. A.A. Zhiglyavskii, *Mathematical Theory of Global Random Search* (Leningrad State University Publishing House, Leningrad, 1985), 296 pp.