number of initial data. Taking into account the above said the conclusion may be made on the fact that application of spatial indexing algorithms is obligatory while developing geofield recovery methods.

The given approaches were successfully used in program package SurfMapper developed by the authors.

*The work was implemented with the support of RFBR grant № 06-07-89190.*

**REFERENCES**

1. Aronov V.I. Methods of constructing maps of geological-geophysical features and geometrization of oil and gas pool on computer. – Moscow: Nedra, 1990. – 301 p.
2. Kanevskiy M.F., Demyanov V.V. Introduction into methods of analyzing environmental data // Problems of environmental and natural resources. – 1999. – № 11. – P. 4–11.
3. Kovin R.V., Markov N.G. Geoinformation technologies for analyzing two-dimensional geofields. – Tomsk: Tomsk University Press, 2006. – 166 p.
4. Burrough P.A., McDonnell R.A. Principles of Geographical Information Systems. – Oxford: Oxford University Press, 1998. – 333 p.
5. Gousie M.B., Williams G., Agnitti T., Doolittle N. CompSurf: An Environment for Exploring Surface Reconstruction Methods on a Grid // Computers & Geosciences. – 2003. – № 9. – P. 1165–1173.
6. Laslo M. Computational geometry and computer graphics on C++ / Translated from English. – Moscow: «BINOM» Press, 1997. – 304 p.
7. Preparata F., Sheimos M. Computational geometry: Introduction. – Moscow: Mir, 1989. – 478 p.

*Received on 16.10.2007*

UDC 504.064(4)

# APPLICATION OF JOHNSON DISRIBUTION TO THE PROBLEM OF AEROSPACE IMAGES CLASSIFICATION

Yu.B. Burkatovskaya, N.G. Markov, A.S. Morozov, A.P. Serykh*

Tomsk Polytechnic University
*Tomsk State University
E-mail: julya@mail.tsu.ru

*Solving the problem of aerospace images classification it was suggested to approximate distribution density of image characteristics by Johnson distribution. The possibilities of such approach were investigated and its availability was shown.*

## Introduction

Aerospace images (AI) are an important source of data on Earth surface state and used, for example, for fire recognition, estimation of territory ecological state etc. Solving the problem of AI automated interpretation is mainly connected with the development of new methods, algorithms and software tools for such image classification [1–3]. The most widespread approach in this case is application of the theory of image statistic recognition [4] using various classifiers. Classification of earth surface types on AI may be carried out using both parametric and nonparametric classifiers. However, the most part of these classifiers do not satisfy increasing requirements to accuracy and speed of classification therefore, the task of development of new high-performance methods and approaches to AI classification is urgent.

In this paper the approach allowing solving the problem of AI classification using Johnson distribution is developed. The results of studying such approach efficiency are given.

## Problem of aerospace image classification

In general case AI consists of $N$ points and includes $K$ data channels. According to electromagnetic spectrum region used for AI formation all AI are divided into ultraviolet, optical, infrared and radar [1]. At present panchromatic (they are also called single-channel) AI are used for analyzing earth surface dynamics. They include one more or less wide spectral region as well as multisensor images (called multi-channel or multiregion) including several layers (channels) obtained simultaneously in various narrow spectrum ranges. Original AI are represented both in widespread formats (for example, Windows Bitmap) and in special formats of geoinformation systems (for example, ER Storage, GeoTiff, Imagine Image) for further processing and interpretation.

Let this or that method is fit for determining that studied AI has $M$ types (classes) of earth surface such as water, vegetation, coniferous forest etc. A sample (AI pixel group) called learning corresponds to each class. Subject to this it is necessary to solve the problem of AI classification that is to refer each point (pixel) of an image to one of classes $M$.

Parametric approach (it is based on hypothesis of normal distribution of feature values) is often used at AI classification with training or methods of nonparametric statistics are used within nonparametric approach or neuronet classification. In the first approach kernel rate

is constructed for feature distribution density and in the second approach the question about belonging of studied AI point to this or that class is solved by nonparametric estimation of features [1]. The first approach gives good results if feature distribution is really close to normal that does not always occur. The second approach does not require knowledge on type of distribution and gives, as a rule, more exact results but possesses at the same time high computational complexity.

Taking into account the above said the authors suggest approximating feature distribution densities by Johnson distributions which depend on three or four parameters and describe adequately wider class of densities than normal distribution depending on two parameters. Besides, random quantities with distributions from Johnson family may be converted into standard normal random quantities that allows using the results obtained for normal distribution for AI classification. The proposed approach is essentially parametric and it means that it can allow increasing classification rate in comparison with nonparametric and neuronet classification methods.

Let us firstly examine the peculiarities of Johnson distributions in respect to the assigned task and obtain the expression for multidimensional distribution density. Then we obtain formula for estimating distribution parameters.

**Johnson distribution (multidimensional case)**

There is multi-channel AI containing $N$ points. Each point is described by vector of $m$ features (both spectral and textural features are included in the set). The problem of finding experimental distribution describing feature distribution density of each image point is set.

N.L. Johnson suggested finding experimental distributions by converting standard normal random quantity [6]. Let $\xi$ be the random value for which Johnson distribution should be selected. In general case the conversion has the form

$$\zeta = \gamma + \eta\tau(\xi;\varepsilon,\lambda);$$
$$-\infty < \gamma < +\infty, \eta > 0, -\infty < \varepsilon < +\infty, \lambda > 0; \qquad (1)$$

where $\tau(\cdot)$ is the certain function; $\gamma$, $\eta$, $\varepsilon$ and $\lambda$ are the distribution parameters and $\zeta$ is the random value distributed by normal law. Johnson suggested the following three various forms of families of functions $\tau(\cdot)$ called respectively the families of Johnson distributions $S_L$, $S_B$ and $S_U$ [6]:

$$S_L: \quad \tau_1(\xi;\varepsilon,\lambda) = \ln\left(\frac{\xi-\varepsilon}{\lambda}\right), \qquad \xi \ge \varepsilon;$$

$$S_B: \quad \tau_2(\xi;\varepsilon,\lambda) = \ln\left(\frac{\xi-\varepsilon}{\lambda+\varepsilon-x}\right), \quad \varepsilon \le \xi \le \lambda+\varepsilon;$$

$$S_U: \quad \tau_3(\xi;\varepsilon,\lambda) = \mathrm{Arsh}\left(\frac{\xi-\varepsilon}{\lambda}\right), \quad -\infty < \xi < +\infty. \qquad (2)$$

As it follows from (2) function $\tau_1(\cdot)$ is fit for describing densities with a carrier restricted on the left, function $\tau_2(\cdot)$ with finite carrier, $\tau_3(\cdot)$ with infinite carrier. However, the above said is not obligatory as the form of the function $\tau(\cdot)$ may depend not only on a carrier but also on the other characteristics of distribution bar chart.

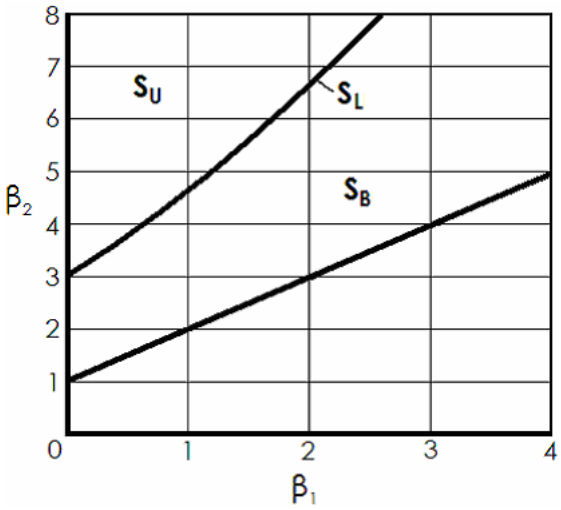


***Fig. 1***. *Diagram for selection of Johnson distribution family*

In [6] it was suggested to select family of Johnson distribution on the basis of normalized indices of skewness $\beta_1$ and excess $\beta_2$ of empirical distribution. Theses indices are determined by the following formula:

$$\sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}}; \quad \beta_2 = \frac{\mu_4}{(\mu_2)^2};$$

where $\mu_i$ is the $i$ central moment of random quantity $\xi$.

Family of distributions $S_L$ is represented in the coordinate system $\beta_1$, $\beta_2$ of the curve as it has a single shape parameter in comparison with families $S_B$ and $S_U$ having two parameters and represented in the form of regions (Fig. 1). The curve corresponding to the family $S_L$ is specified according to [6].

Let us examine a random vector $\xi=[\xi_1,...\xi_m]^T$. Let us assume that its component with number $i$ follows Johnson distribution with function $\tau_{ji}(\cdot)$ and parameters $(\gamma_i,\eta_i,\varepsilon_i,\lambda_i)$. Applying Johnson transformation (1) the random vector $\zeta=[\zeta_1,...\zeta_m]^T$ is obtained

$$\zeta = \gamma + \eta\tau(\xi;\varepsilon,\lambda), \qquad (3)$$

where

$$\gamma = [\gamma_1,\gamma_2,...,\gamma_m]^T, \ \varepsilon = [\varepsilon_1,\varepsilon_2,...,\varepsilon_m]^T, \ \lambda = [\lambda_1,\lambda_2,...,\lambda_m]^T;$$

$$\eta = \begin{bmatrix} \eta_1 & 0 & ... & 0 \\ 0 & \eta_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \eta_m \end{bmatrix}, \quad \tau(\xi;\varepsilon,\lambda) = \begin{bmatrix} \tau_{j1}(\xi_1;\varepsilon_1,\lambda_1) \\ \tau_{j2}(\xi_2;\varepsilon_2,\lambda_2) \\ ... \\ \tau_{jm}(\xi_m;\varepsilon_m,\lambda_m) \end{bmatrix},$$

$$j_i \in \{1,2,3\}.$$

Components of vector $\xi$ and vector $\zeta$ are more often dependent in practice. Separate components of vector $\zeta$ follow the standard normal distribution. It does not mean, of course, that vector $\zeta$ has multinormal distribution. However, the cases when joint distribution of Gaussian random variables are not Gaussian ones are not often in practice. Therefore, let us assume that joint distribution of components of vector $\zeta$ is Gaussian as well. Let us examine the matrix covariations.

$$\Sigma = \begin{bmatrix} 1 & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{12} & 1 & \dots & \sigma_{2m} \\ \dots & \dots & \dots & \dots \\ \sigma_{1m} & \sigma_{2m} & \dots & 1 \end{bmatrix}, \quad \sigma_{ij} = M\{\zeta_i \zeta_j\}. \qquad (4)$$

Then, distribution density of vector $\zeta$ has the form

$$f_\zeta(y) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \exp\{-\frac{1}{2} y^T \Sigma^{-1} y\}. \qquad (5)$$

Construction of covariaton matrix (4) and use of formula (5) allows taking into account the linear connections between components of vector $\zeta$.

Now let us obtain density of vector $\xi$. Let us introduce matrix $\tau'(\xi;\varepsilon,\lambda)$ the elements of the main diagonal of which are derivative functions $\tau_{ji}(\xi_i;\varepsilon_i,\lambda_i)$:

$$\tau'(\xi;\varepsilon,\lambda) = \begin{bmatrix} \tau'_{j1}(\xi_1;\varepsilon_1,\lambda_1) & 0 & \dots & 0 \\ 0 & \tau'_{j2}(\xi_2;\varepsilon_2,\lambda_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tau'_{jm}(\xi_m;\varepsilon_m,\lambda_m) \end{bmatrix}.$$

Taking into account (3) we have

$$f_\xi(x) = |\eta||\tau'(x;\varepsilon,\lambda)| \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \times$$

$$\times \exp\{-\frac{1}{2}[\gamma + \eta\tau(x;\varepsilon,\lambda)]^T \Sigma^{-1} [\gamma + \eta\tau(x;\varepsilon,\lambda)]\}.$$

Let us introduce notations, $\delta = \eta^{-1}\gamma$, $D = \eta^{-1}\Sigma\eta^{-1}$. Finally we have

$$f_\xi(x) = |\tau'(x;\varepsilon,\lambda)| \frac{1}{\sqrt{(2\pi)^m |D|}} \times$$

$$\times \exp\{-\frac{1}{2}[\tau(x;\varepsilon,\lambda) - \delta]^T D^{-1} [\tau(x;\varepsilon,\lambda) - \delta]\}. \qquad (6)$$

If multivariate standard density of distribution with parameters $(\delta, D)$ is denoted by $\varphi(x,\delta,D)$ then formula (6) takes on form

$$f_\xi(x) = |\tau'(x;\varepsilon,\lambda)| \varphi(\tau(x;\varepsilon,\lambda);\delta, D).$$

Owing to the fact that Johnson transformation of random variables result in standard Gaussian variables the results obtained for normal random variables may be used for operation with Johnson distributions. Let us consider, in particular, estimation of the parameters of multivariate density of distribution.

Let us firstly, state the results obtained in [6] for one-dimensional random variables. There is a set of independent random variables $\{x_1,...,x_N\}$ having Johnson distribution. Let us notice that parameters $\varepsilon$, $\lambda$ make a sense of parameters of random variable scale. As the point brightness which may be changed in a certain finite interval of values (often for 0 to 255) is used as an observed random variable in the tasks of AI processing and interpretation then the interval of random variable change is known a priori and this information may be used when estimating parameters $\varepsilon,\lambda$.

For distribution of $S_L$ where

$$\zeta = \gamma + \eta \ln\left(\frac{\xi - \varepsilon}{\lambda}\right),$$

the variables are changed and a number of parameters is decreased to three:

$$\zeta = \gamma^* + \eta \ln(\xi - \varepsilon), \quad \gamma^* = \gamma - \eta \ln\lambda.$$

Parameter $\varepsilon$ here makes sense of left distribution limit and it may be taken as equal to zero or minimal value of feature in learning sample in the task of AI identification. At known $\varepsilon$ estimations for parameters $\gamma^*$ and $\eta$ have the form

$$\hat{\eta} = \frac{1}{\widehat{\sigma}}, \quad \hat{\gamma}^* = -\frac{\widehat{\mu}}{\widehat{\sigma}},$$

where

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \ln(x_i - \varepsilon), \quad \widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N} \ln(x_i - \varepsilon)^2}{N} - \widehat{\mu}^2}.$$

For the family of distributions $S_B$ in the case when parameters $\varepsilon$ (left distribution limit) and $\lambda$ (width of distribution carrier) are known, parameters $\eta$ and $\gamma$ are estimated by comparison of two empirical percentiles to two appropriate percentiles of standard normal distribution:

$$\hat{\eta} = \frac{z_{1-\alpha'} - z_\alpha}{\ln\left(\frac{(x_{1-\alpha'} - \varepsilon)(\varepsilon + \lambda - x_\alpha)}{(x_\alpha - \varepsilon)(\varepsilon + \lambda - x_{1-\alpha'})}\right)},$$

$$\hat{\gamma} = z_{1-\alpha'} - \hat{\eta} \ln\left(\frac{x_{1-\alpha'} - \varepsilon}{\varepsilon + \lambda - x_{1-\alpha'}}\right),$$

where $z_\alpha$ and $z_{1-\alpha'}$ represent $\alpha 100$ and $(1-\alpha')100$ percentiles of standard normal distribution and $x_\alpha$ and $x_{1-\alpha'}$ are the corresponding empirical percentiles. If parameters $\varepsilon$ and $\lambda$ are unknown then they may be taken equal to feature minimal value in learning sample and difference between maximal and minimal value of the feature

The questions of estimating parameters of one-dimensional random variable following Johnson distribution with unknown limits of change are examined in detail in [6]. The method of estimating distribution parameters of the family $S_U$ is based on numerical solution of nonlinear equation and examined in [7].

Now let $\{x_1,...,x_N\}$ are the $m$-dimensional vectors the components of which follow Johnson distribution that is $x_i = [\xi_1^i,...,\xi_m^i]^T$. Let us transform vectors $x_i$ into vectors $y_i$ similarly to (3) but their estimations are used instead of parameters.

$$y_i = \hat{\gamma} + \hat{\eta}(x_i;\hat{\varepsilon},\hat{\lambda});$$

$$\hat{\eta} = \begin{bmatrix} \hat{\eta}_1 & 0 & \dots & 0 \\ 0 & \hat{\eta}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{\eta}_m \end{bmatrix}, \quad \hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_m \\ \dots \\ \hat{\gamma}_m \end{bmatrix}. \qquad (7)$$

Then let us estimate parameters $(\delta, D)$ of distribution density (6)

$$\hat{\delta} = -\hat{\eta}^{-1}\hat{\gamma};$$

$$\widehat{\sigma}_{ij} = \frac{1}{N}\sum_{l=1}^{N} y_i^l y_j^l, \quad \widehat{\sigma}_{ii} = 1;$$

$$\widehat{D} = \hat{\eta}^{-1}\hat{\Sigma}\hat{\eta}^{-1}. \qquad (8)$$

Substituting the obtained estimations into (6) we obtain the estimation of joint distribution density of random vector components.

#### Results of AI classification

Taking into account the obtained theoretical results the generalized algorithm of AI classification by maximum likelihood method with Johnson distributions looks like this.

*Stage 1.* To select the family of Johnson distributions and obtain estimation of parameters of one-dimensional Johnson distribution for each feature of each class (of learning sample).

*Stage 2.* to transform vectors of pixel features of learning samples by the formula (3) to vectors with standard normal distribution using parameters obtained at the stage 1 and obtain estimations of covariation and parameters of multivariate distribution for each class by the formulas (4), (7), (8).

*Stage 3.* To calculate for each AI pixel the estimation of feature distribution density by the formula (6) for each class using parameters obtained at the stage 2.

*Stage 4.* To refer each AI pixel to the class returning maximal value of feature distribution density.

To study the efficiency of suggested approach allowing solving the problem of AI classification using Johnson distribution the appropriate software is developed taking into account the given algorithm.

To classify the images the real AI of different detail obtained on space systems of earth remote sensing were used in the experiments. Kappa-index of agreement (KIA) and cumulative probability of correct classification (PCC) were used as accuracy indices of classification [1]. As an example, the results of classification of the original multi-channel space photograph (seven spectral zones) of coastal zone fragment in San Diego (Fig. 2, *a*) made by the system Landsat TM (USA) are given in Fig. 2 [8]. A number of classes for learning according to the results of the paper [2] is chosen to be equal ten.

For visual estimation of AI classification results the image obtained at classification with Johnson distribution is given in Fig. 2, *c*, and the result of classification with normal feature distribution is given in Fig. 2, *b*. It is seen that at classification with normal distribution the breakers line in the original photo (Fig. 2, *a*, left low corner) not included into a number of training classes was referred to the class «concrete constructions» (Fig. 2, *b*) whereas at classification with Johnson distribution these pixels were not classified as distant rather far in a space of features from all used classes (black pixels in Fig. 2, *c*).

a

b

c

***Fig. 2.*** *Examples of AI classification*

Quantitative indices of AI classification accuracy for the described example are given in the Table. It may be noted that classification with Johnson distribution is performed with rather low accuracy than classification

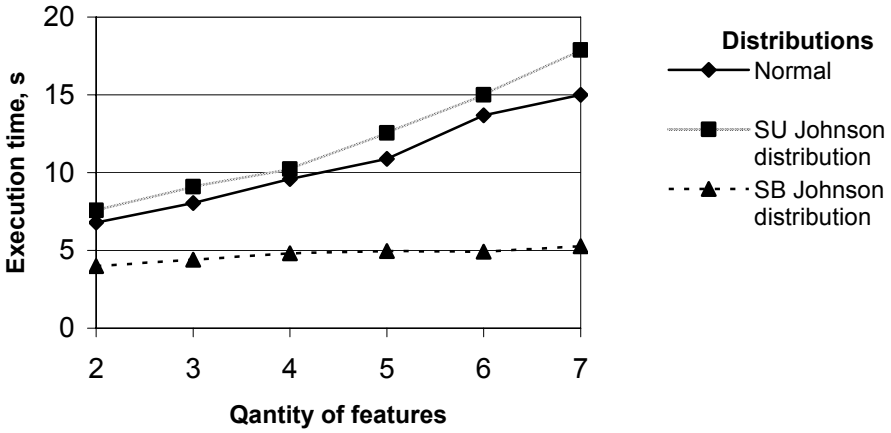


***Fig. 3.*** *Dependence of classification time on quantity of features*

in the case of assumption about normal distribution of features (the given image is practically an ideal case when the distribution is close to normal). Besides, because of boundedness of distribution carrier of the family $S_B$ Johnson (this very family was selected for the majority of features of the majority of classes taking into account values $\beta_1$ and $\beta_2$ and Fig. 1) 5,1 % of pixels were not classified.

***Table.*** *Classification accuracy, %*

| Type of density estimation | KIA | PCC |
|---|---|---|
| Normal distribution | 97,82 | 98,61 |
| Johnson distribution | 96,91 | 98,19 |

The results of studying computing efficiency of the proposed approach are given in Fig. 3 (image size is 500×500 pixels). The expenditures of computer time in the case of AI classification with normal feature distribution are shown here as well for comparison. It is seen that the rate of performance of algorithm implementing the suggested approach to AI classification is higher than in the case of classification with normal distribution. This result is explained by the boundedness of distribution carrier of the family $S_B$ Johnson as at class search, those of them in definitional domain of which the given pixel is not found, by one feature at least, are rejected already at the initial stage of computation. The rate of algorithm operation with compulsory use of distribution family $S_U$ which has infinite carrier is rather lower than a traditional method.

## Conclusion

To solve the problem of aerospace image classification it was suggested to approximate densities of image feature distribution by Johnson distributions. Derivation of the formula for multivariate density of feature distribution and estimation of parameters of their distribution density is given.

A set of computer experiments with real aerospace images was carried out for studying the efficiency of suggested approach. The first encouraging results indicating the availability of Johnson distribution use when solving the problem of classification with learning were obtained.

*The work is supported by the grant RFBR № 06-07-89190a.*

## REFERENCES

1. Zamyatin A.V., Markov N.G. Analysis of earth surface dynamics by the data of earth remote sensing. – Moscow: Fizmatlit, 2007. – 176 p.
2. Giacinto G., Roli F. Dynamic Classifier Selection Based on Multiple Classifier Behavior // Pattern Recognition. – 2001. – V. 34. – № 9. – P. 179–181.
3. Podest E., Saatchi S. Application on Multiscale Texture in Classifying JERS-1 Radar Data Over Tropical Vegetation // International Journal of Remote Sensing. – 2002. – V. 23. – № 7. – P. 1487–1506.
4. Fukunaga K. Introduction into statistical theory of image identification. – Moscow: Nauka, 1979. – 368 p.
5. Haralick R.M., Shapiro L.G. Computer and Robot Vision. – Massachusetts: Addison-Wesley, 1992. – P. 199–218.
6. Khan G., Shapiro S. Statistical models in engineering problems. – Moscow: Nauka, 1969. – 308 p.
7. Johnson N.L. Tables to Facilitate Fitting SV Frequency Curves. – Oxford: Biometrika Trust, 1965. – P. 52–57.
8. http://www.leica-geosystems.com.

*Received on 11.10.2007*