Ukrainian Journal of Remote Sensing
journal homepage: www.ujrs.org.ua

https://doi.org/10.36023/ujrs.2026.13.3.313

UDC 004.932:621.391:528.8

Analysis of the Impact of Lossy Compression of Multichannel Space Images on the Informativeness of Spectral Features

S. Y. Shulhin*, https://orcid.org/0009-0005-3997-3998
I. K. Vasylieva, http://orcid.org/0000-0002-1378-1104
Department of Information and Communication Technologies, National Aerospace University, 61070 Kharkiv, Ukraine

The development of remote sensing technologies and the high spatial resolution of modern satellite systems have led to an increase in the volume of observation data. The problem of storing and transmitting large volumes of data necessitates the use of compression methods. However, lossy compression inevitably leads to distortions of the statistical structure of the spectral components of data, which affects the results of their subsequent analysis and interpretation. This article presents a study of the influence of the most common compression technologies (JPEG, JPEG2000, AVIF, WebP) on the informativeness of spectral features in the statistical classification of land surface types based on Sentinel-2 data. To describe the empirical distributions of the features, the Johnson S_B-distribution is used, taking into account the non-Gaussian nature of the distributions and the high level of cross-correlation between the R, G, and B channels. A quantitative assessment of the feature informativeness is based on local and generalized divergence criteria. An analysis of images with varying structural complexity revealed the nonlinear nature of the deformation of the statistical moments of reference classes under the influence of lossy compression. For images with a relatively simple structure, coding algorithms shift the distribution mode and reduce within-class scatter, which can lead to increased generalized divergence. Conversely, for images with high spatial detail, even minimal compression causes a significant decrease in the feature informativeness due to the reduced between-class distances. These results highlight the need to develop adaptive classifiers for compressed images that can dynamically compensate for codec-specific statistical errors.

Keywords: lossy compression, multichannel space images, informativeness, generalized divergence, Johnson S_B-distribution, statistical analysis.

© S. Y. Shulhin, I. K. Vasylieva. 2026

Introduction

Effective solutions for automatic recognition of objects on the Earth's surface using remote sensing data directly depend on the choice and discriminative capabilities of the feature space. The high spatial resolution of sensors in modern space systems, such as Sentinel-2, and the ability to observe across a wide spectral range lead to a significant increase in the volume of information, necessitating the use of data compression algorithms to optimize the use of communication channels and storage systems. Along with the traditional JPEG standard, modern high-performance codecs such as JPEG2000, WebP (Barina, 2021), and AVIF (Corino et al., 2026) are attracting increasing attention and are considered promising alternatives for storing multidimensional visual data. However, any lossy compression inevitably distorts the statistical structure of the spectral components of an image. Removing high-frequency components and quantizing transform coefficients alter the conditional probability distributions of features in the reference regions used for training and testing classifiers. This makes it relevant to carry out a quantitative assessment, as well as a statistical analysis of how compression processes carried out by different types of codecs affect the information content of spectral characteristics used to classify objects.

The impact of image compression on the stochastic properties of spectral features of observed objects, as well as on the classification accuracy of compressed images, has been studied in many works. In particular, the paper (Gandor and Nalepa, 2022) presents the results of a study on the impact of JPEG compression on the performance of object detectors based on convolutional neural networks; it is shown that with an increase in the degree of compression, the recall metric decreases, while the precision value remains practically the same. The work (Pomarico et al., 2022) proposed a method for quantifying the statistical distortions of images caused by JPEG compression; it is shown that 10:1 compression together with a decrease in bit depth from 16 to 8 leads to a change in the predictions obtained using supervised learning models by more than 10 standard deviations. Thus, the use of compressed data sets leads to an increase in the pre-dictive uncertainty of decisions in .

*Corresponding author / Автор для кореспонденції: S. Y. Shulhin / С. Ю. Шульгін / s.y.shulhin@student.khai.edu

This is an Open Access article under the CC BY licenses (http://creativecommons.org/licenses/by/4.0/)
Стаття опублікована на умовах відкритого доступу за ліцензією CC BY (http://creativecommons.org/licenses/by/4.0/)

diagnostic applications. The paper (Sultana et al., 2023) proposes a solution to the problem of an efficient method for searching compressed images based on their content; the results of a study on the impact of lossy compression on the color and texture characteristics of images and, accordingly, on the performance of the search system are presented.

The paper (Abramova et al., 2022) describes a methodology for statistical and spatial-spectral analysis of distortions introduced during lossy compression and presents the results of such analysis for remote sensing images compressed by the AGU coder. These results prove that the distortions under study are uncorrelated, but are statistically characterized by spatial variability (more intense distortions are inherent in locally active areas).

In the work (Kovalenko and Lukin, 2023), stochastic characteristics of compression distortions caused by the BPG coder were determined: both functional (histograms) and numerical (kurtosis, asymmetry, mean, variance); it has been shown that under certain conditions, the distortion distribution can be considered close to Gaussian, but this distribution can also have a «heavy tail». The main factors influencing the type of compression distortion distribution include the presence of noise in the original image, the degree of complexity of the image structure, and the value of the quality parameter Q, which is directly related to the compression ratio (CR).

The paper (Shulhin and Vasylieva, 2025) presents the results of a comparative statistical analysis of difference images (defined as the difference between the original and compressed images). Two types of test images were used for the study: synthetic and real (obtained from the Sentinel-2 satellite), to which various compression algorithms (JPEG, JPEG2000, AVIF and WebP) were applied with changing the values of the Q parameter in the range from 90 to 30. It is shown that a decrease in the parameter Q (an increase in CR) leads to an increase in the dispersion and excess of compression distortion, while the values of the mean and skewness are close to zero. Accordingly, the distribution of compression distortions, initially (at Q = 90) close to normal, remains quite symmetrical with increasing CR, but acquires a «heavy» exponential «tail» (i.e., becomes close to the Laplace distribution).

Thus, lossy image compression inevitably leads to changes in pixel values. Although these changes may be visually imperceptible, they can lead to significant changes in the stochastic characteristics of the spectral features of objects in the compressed image. In other words, compression may lead to changes in the type and/or distribution parameters of spectral features of classes, which affects any image processing methods that use statistical information, including adaptive spatial filtering methods, statistical segmentation and classification methods.

The aim of this work is to study the impact of lossy image compression procedures (JPEG, JPEG2000, AVIF, WebP) on the informativeness of spectral features in land surface classification using Sentinel-2 data. Generalized divergence (Vasil'eva and Popov, 2016) is used as the informativeness criterion.

1. Defining a model for spectral features

In real-world image recognition problems, features (in particular, spectral ones) are never deterministic, so taking into account the stochastic nature of features allows us to construct their physically adequate models for making informed decisions about the class of the observation object. There are many reasons to consider the spectral image of an object not as a point in feature space, but as a joint probability distribution of random feature values. First, the reflectivity of objects in the spectral ranges under study is unstable over time; it depends on illumination, atmospheric conditions, and other factors. Second, noise effects (hardware noise in the sensor system, noise in communication channels during measurement data transmission, noise during image digitization), optical effects, perspective distortions, and the effects of sudden changes in illumination exacerbate this uncertainty. Therefore, observation results are realizations of random variables, and the use of reliable stochastic information about the properties of objects improves the accuracy of their recognition. Moreover, taking into account the stochastic nature of the data allows us to move from rigid decision rules to more flexible probabilistic models: a model trained on distributions, rather than on reference vectors, works more stably with new data that was not in the training set.

In accordance with the statistical approach to constructing classification procedures, conditional probability density functions (PDFs) $f(\vec{x};\vec{\theta}|a_k)$ are used to represent reference class descriptions, where $\vec{x}$ is the vector of features, $\vec{\theta}$ is the vector of distribution parameters, a_k is the class of the object, $k = 1, 2,\ldots K$, where K is the number of classes.

Thus, information about each class of objects is contained in the parameters of the joint distribution of features, which makes it possible to identify decision regions in favor of the kth class G_k in the feature space.

The classifier is considered optimal if the regions $\{G_k\}_{k=1}^{K}$ do not overlap, that is, the following condition must be met: an object s (in the case of pixelwise classification, this is an individual pixel of the image) belongs to the class a_k if the corresponding feature vector $\vec{x}*$ belongs to the region G_k:

$$s \in a_k \Leftrightarrow \vec{x}* \in G_k, k = 1,\ldots,K.$$

In real recognition problems, the regions $\{G_k\}_{k=1}^{K}$ usually overlap. However, the stochastic approach in this case also allows one to find a dividing surface between classes that will be optimal with respect to the chosen criterion (for example, the Bayesian criterion (Duda et al., 2001; Jain et al., 2000) minimizes the average risk associated with the generalized losses due to classification errors).

Since a priori descriptions (reference models) of classes are unknown, either completely or at least in their parameters, their estimates must first be found. As noted above, according to statistical recognition theory, reference models of classes are the joint distribution densities $f(\vec{x};\vec{\theta}|a_k)$ of their features. The histogram

method is most often used to estimate the distribution density of training samples. Clearly, the quality of the classification procedure depends on the adequacy of statistical models. Inadequate and inaccurate distribution density estimates can undermine the effectiveness of statistical decision-making algorithms.

Distributions of real spectral data in the optical range (R, G, B) typically have a pronounced non-Gaussian shape and are characterized by strong cross-correlations with linear correlation coefficients in the range of 0.7–0.98. For an adequate analytical description of such random variables, it is advisable to use the Johnson distribution system $\{S_L, S_B, S_U\}$ (Johnson, 1949), based on nonlinear transformations of the form:

$$z = \gamma + \eta\, \tau(x; \varepsilon, \lambda), \quad (1)$$

where z is a normalized random variable; ε is the bias parameter; λ is the scale parameter; η and γ are shape parameters; $\lambda > 0$, $\eta > 0$.

Since the pixel values of digital satellite images (proportional to the intensity of the received signal in a certain spectral channel) are always limited by a fixed dynamic range, it is advisable to take the Johnson S_B-distribution with a finite support to approximate the intensity distribution of spectral features. In this case, the transformation function has the form

$$\tau(x;\varepsilon,\lambda) = \ln\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right),\ \varepsilon \le x \le \varepsilon+\lambda. \quad (2)$$

The probability density function for the Johnson S_B-distribution is described by the formula

$$f(x) = \frac{\eta\lambda}{\sqrt{2\pi}(x-\varepsilon)(\varepsilon+\lambda-x)} \exp\left[-\frac{1}{2}\left(\gamma+\eta\ln\left(\frac{x-\varepsilon}{\varepsilon+\lambda-x}\right)\right)^2\right]. \quad (3)$$

The Johnson S_B-distribution is a universal statistical model that can approximate practically any unimodal distribution, as well as a wide range of bimodal distributions. The disadvantage of the model (3) is the lack of an explicit analytical relationships between the distribution parameters and statistical moments (mean, variance, skewness and kurtosis). Therefore, to estimate the parameter vector $\vec{\theta} = \{\varepsilon, \lambda, \eta, \gamma\}$, iterative optimization methods are commonly used, aimed at minimizing the total squared error between the empirical histogram and the theoretical model.

Thus, the spectral features under study were described by model (3), the parameters of which for a specific feature depended on the class of the object in the image and were estimated using histograms obtained using training samples.

2. Information measures and the generalized divergence criterion

Informativeness characterizes the distinctive properties of a feature, its effectiveness in dividing the observed set of objects into non-overlapping sets of classes.

In statistical recognition, criteria based on either feature metrics for different classes or recognition accuracy estimates are used to compare the informativeness of features. One of the most common criteria is the statistical probability of misclassification (and the associated risk of loss). Theoretically, the probabilities of erroneous decisions

$$P_{uv} = P\left\{\vec{x}^* \in G_u \,|\, s \in a_v\right\} = \int_{G_u} f(\vec{x}|a_v)d\vec{x},\ u \neq v,\ u,v = 1...K$$

quantitatively express the volumes of those regions of intersection of PDFs (limited by the decision-making surfaces) for which a decision in favor of the correct class a_u is less likely than alternative hypotheses. In this case, the transformation of the probability distribution due to distortions due to lossy compression should lead to a change in the recognition error probabilities. However, the probability criterion for misclassification is related to the chosen decision criterion, so it rather characterizes the effectiveness of the recognition system rather than the informativeness of the classification features.

If feature evaluation is performed independently of classification accuracy, then informativeness indicators are based either on distance metrics between samples from different classes (e.g., Euclidean, Mahalanobis distances, within-class and between-class scattering matrices used in discriminant analysis (Duda et al., 2001)) or on integral characteristics of PDFs of class features (Kullback divergence (Kullback and Leibler, 1951)).

For comparative analysis of the discriminant properties of features without reference to a specific classifier type, it is convenient to use the Kullback divergence (or classical relative entropy), which is directly related to the logarithm of the likelihood ratio:

$$J(k, j) = \int_{-\infty}^{\infty}\left[f(x|a_k) - f(x|a_j)\right]\ln\frac{f(x|a_k)}{f(x|a_j)}dx. \quad (4)$$

Divergence (4) is a measure of the relative importance of a feature for classification: the greater the divergence, the more useful the information that the feature provides for distinguishing between classes a_k and a_j. At this, if $k = j$, then $J(k, j) = 0$, otherwise $J(k, j) > 0$. However, formula (4) is applicable only in the case of two classes of objects.

In the case of multiple-alternative recognition ($K > 2$), the measure of distinguishability of the kth class among all other classes is local divergence (Vasil'eva and Popov, 2016):

$$J(k) = \int_{-\infty}^{\infty}\left[f(x|a_k) - f(x|\tilde{a}_k)\right]\ln\frac{f(x|a_k)}{f(x|\tilde{a}_k)}dx, \quad (5)$$

where $f(x|\tilde{a}_k)$ is the mixture of probability densities for all alternative classes except the class a_k:

$$f(x|\tilde{a}_k) = \frac{1}{1-P(a_k)}\sum_{i\neq k} P(a_i)\, f(x|a_i). \quad (6)$$

where $P(a_i)$ is a prior probability of the corresponding class a_i, $i = 1\ldots K$.

Local divergence (5) is a measure of how much one class a_k differs from all others $\{a_i\}_{i\neq k}^{K-1}$ in terms of some feature x (or set of features $\vec{x}$).

The average value of local divergences determines the degree of distinguishability of the entire set of classes according to the selected set of features and is

called the generalized divergence (Vasil'eva and Popov, 2016):

$$J_x = \sum_{k=1}^{K} P(a_k) J(k). \quad (7)$$

Generalized divergence (7) can be used to compare the informativeness of different subsets of a set of classification features.

A set of features is considered sufficiently informative for the kth class if $J(k) \geq \Theta_k$, where Θ_k is the minimum sufficient divergence (minimum sufficient distinguishability) (Vasil'eva and Popov, 2016):

$$\Theta_k = \left[\frac{1-P(a_k)^2}{P(a_k)^2}\right] \ln \frac{[1-P(a_k)]^2}{P(a_k)^2}.$$

The value of Θ_k depends on the number of classes and on the a priori probability of the appearance of objects of class a_k in the observation sample. For the case of four equally probable classes ($P(a_k) = 1/K = 0.25$) the value of the minimally sufficient divergence is $\Theta = 17.578$.

Thus, local (5) and generalized divergences (7) are indicators of statistical separability of classes and can be used in procedures for selecting the most informative classification features according to the criterion of maximum divergence.

3. Experimental studies and analysis of results

The study was conducted to investigate the influence of lossy compression on the information content of spectral characteristics of classes in two multichannel optical images with different spatial structure complexity. These test images were obtained from the Sentinel-2 satellite (Lukin et al., 2020; Lukin et al., 2022). Four generalized classes of objects were distinguished in each image: 1 – Urban, 2 – Water, 3 – Vegetation and 4 – Bare soil.

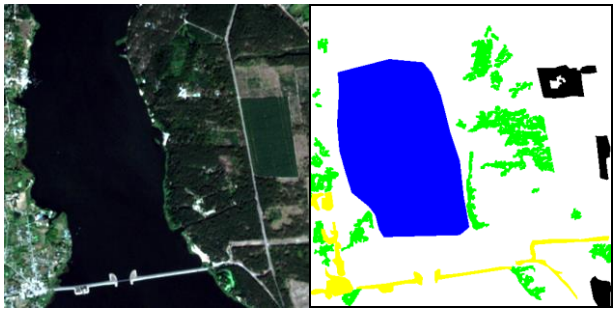
Supervised learning methods were used to solve the research problems. Based on actual data on the Earth's surface (Kharkiv region, Ukraine), test areas were defined that represented relatively homogeneous zones corresponding to a separate class. The resulting areas (color masks) were marked with conditional colors indicating the corresponding class: 1 – yellow, 2 – blue, 3 – green, 4 – black. The test images and class maps are shown in Figs. 1 and 2.

Test images were compressed using the JPEG, JPEG2000, WebP, and AVIF codecs; the compressed image quality parameter Q varied from 90 (minimum compression) to 30 (maximum compression).

Reference areas corresponding to individual object classes were identified in the test images (original and compressed) using masks (Fig. 1, b, Fig. 2, b), and training samples of the classes were formed. The statistical analysis of the training samples consisted of constructing histograms of the features R, G and B for four classes and estimating the parameters of the approximation model (3).

For example, Fig. 3 shows the histograms of feature B for the Vegetation class on the test image S1 (original and compressed, Q = 30) and Johnson S_B density distribution plots, which approximate the empirical distributions.

Reference descriptions (3) of classes based on feature B for test images S1 and S4 and their transformation as a result of compression by the AVIF codec with parameter Q = 30 are presented in Figs. 4 and 5, respectively.



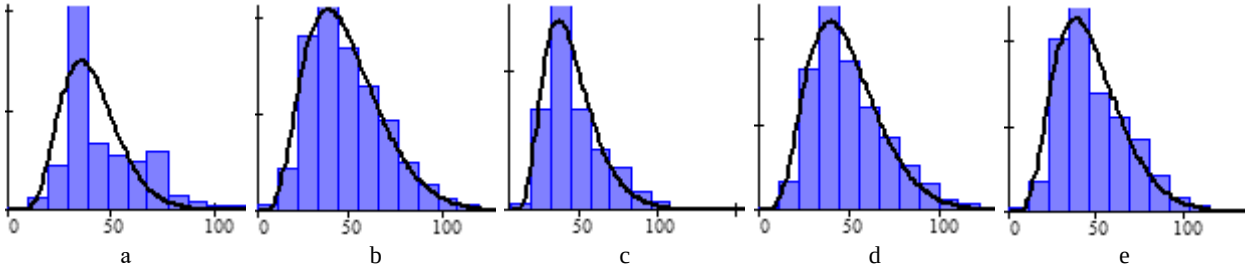
a b

Fig. 1. Test multichannel image S1 (a) and color masks of reference areas (b)



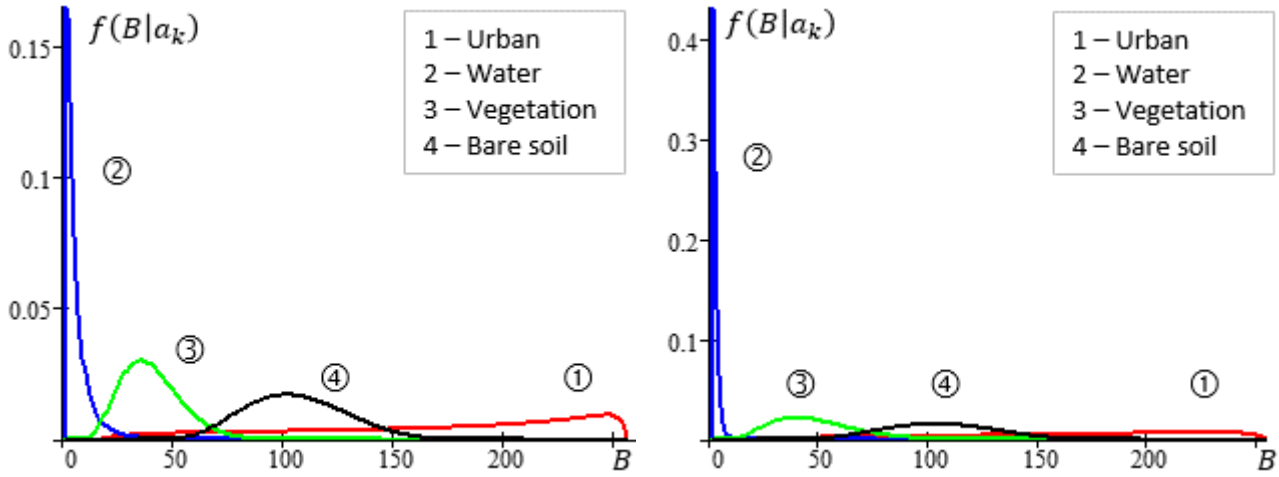
a b

Fig. 2. Test multichannel image S4 (a) and color masks of reference areas (b)



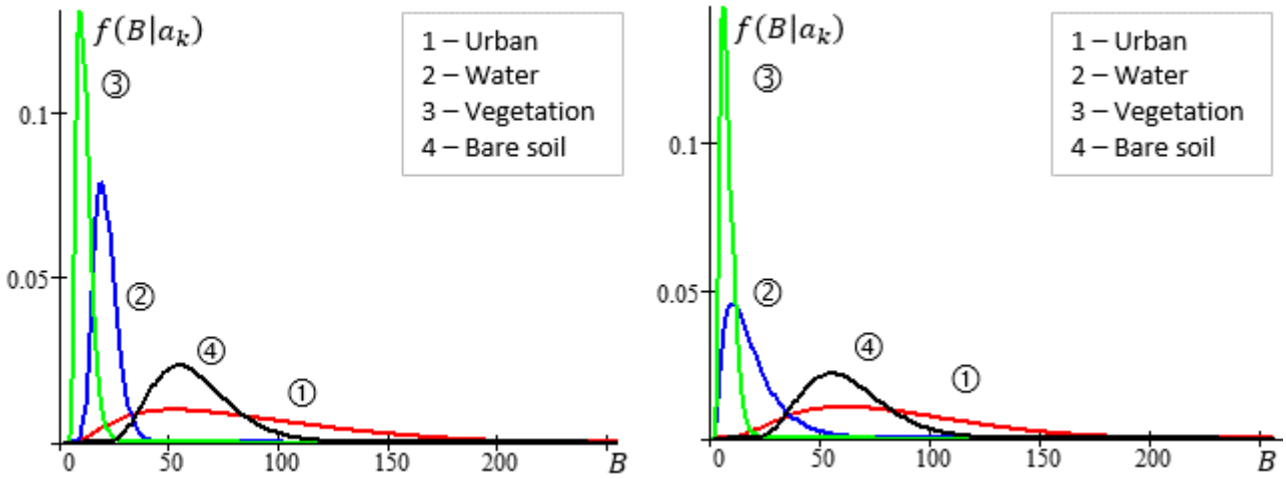
a b c d e

Fig. 3. Histograms of feature B for the Vegetation class in image S1 and their approximation by model (3): a – without compression, b – JPEG (Q = 30), c – JPEG2000 (Q = 30), d – AVIF (Q = 30), e – WebP (Q = 30)



a b

Fig. 4. Feature B statistical models (3) for four classes of objects in image S1:
a – without compression, b – AVIF (Q = 30)



a b

Fig. 5. Feature B statistical models (3) for four classes of objects in image S4:
a – without compression, b – AVIF (Q = 30)

3.1. Analysis of spectral feature informativeness for the test image S1 of simple structure

At the first stage of the study, the informativeness (discriminatory ability) of the spectral features R, G and B in the image S1, characterized by the presence of large objects with statistically homogeneous spectral properties, was assessed. The compression effect of various codecs was studied at a fixed compression level (Q = 30); the results of calculating local $J(k)$ (5) and generalized J_x (7) divergences, $x \in \{R, G, B\}$, are presented in Tables 1–3.

Table 1. Estimates of local $J(k)$ and generalized J_R divergences for channel R

Object class	Local divergence $J(k)$ at Q = 30				
	Uncompressed	JPEG	JPEG2000	AVIF	WebP
Urban	3.631	6.946	9.255	6.616	7.472
Water	12.863	11.892	17.077	12.327	14.163
Vegetation	9.064	7.469	13.032	7.996	9.012
Bare soil	2.795	3.468	3.420	2.836	3.213
Generalized divergence J_R	7.088	7.444	10.696	7.444	8.465

Table 2. Estimates of local $J(k)$ and generalized J_G divergences for channel G

Object class	Local divergence $J(k)$ at Q = 30				
	Uncompressed	JPEG	JPEG2000	AVIF	WebP
Urban	6.152	3.553	8.563	3.529	2.607
Water	10.398	4.036	17.173	6.542	10.179
Vegetation	5.517	3.117	8.944	5.016	5.806
Bare soil	4.416	3.520	3.250	3.354	3.109
Generalized divergence J_G	6.621	3.557	9.483	4.610	5.425

Table 3. Estimates of local $J(k)$ and generalized J_B divergences for channel B

Object class	Local divergence $J(k)$ at Q = 30				
	Uncompressed	JPEG	JPEG2000	AVIF	WebP
Urban	3.436	3.001	5.078	2.664	2.644
Water	21.015	34.238	13.289	29.920	34.172
Vegetation	11.684	10.724	6.944	11.349	12.587
Bare soil	4.943	3.563	4.296	3.552	3.747
Generalized divergence J_B	10.270	12.882	7.402	11.871	13.288

Analysis of the data in Tables 1–3 demonstrates variable behavior of the informativeness of spectral features under lossy compression. In particular, the R feature exhibits an increase in generalized divergence after compression by all codecs, reaching a maximum value of 10.696 for the JPEG2000 codec. Moreover, local divergence values increase not only for large size objects (Water, Vegetation, Bare Soil), but also for the Urban class, which includes narrow, extended objects in satellite images, such as roads, bridges, and dams. Thus, although the distortions resulting from lossy compression are critical for line and point objects due to their small size and hence smaller ratio of object area to distortion area (artifacts), changes in the distribution of spectral characteristics can lead to increased distinguishability of these classes from the rest. Furthermore, in the blue B channel for the Water class, an anomalous increase in local divergence was recorded – from 21.015 to 34.238 for JPEG and WebP compression. This trend is expected and indicates that in images with simple structure, block coding algorithms act as a low-pass filter, smoothing out brightness variations within large homogeneous objects, which leads to a narrowing of the empirical dispersion of features.

3.2. Analysis of spectral feature informativeness for the test image S4 of complex structure

To study the impact of lossy compression on the information content of spectral features in the case of a complex, highly detailed image structure, test image S4 (Fig. 2) was used.

Summary estimates of the informativeness metrics for R, G, and B features for the high-compression conditions (Q = 30) are presented in Table 4.

Table 4. Comparative analysis of generalized divergence J_x for image S4 (Q = 30)

Spectral feature	Uncompressed image	JPEG	JPEG2000	AVIF	WebP
Channel R	14.728	9.390	7.713	12.533	14.507
Channel G	6.901	6.605	7.655	7.702	7.447
Channel B	10.448	7.642	8.255	8.136	8.438

Unlike simpler images, compression of images with a complex spatial structure in most cases leads to the expected reduction in the informativeness of spectral features compared to the original image, which is especially noticeable in the R and B channels. It should be noted that the generalized divergence changes slightly for the G feature: when using the JPEG codec, its value decreases by 4%, and when using AVIF, it increases by approximately 11%, while for features R and B, this criterion only decreases. The largest decrease in the J_x criterion value for feature B is almost 27% (for JPEG); for feature R, the J_x criterion value decreases to the greatest extent – by almost 48% (for JPEG2000).

The obtained results confirm the advantages of modern compression formats (AVIF and WebP) compared to classic JPEG and even the more advanced JPEG2000. Compression errors in JPEG (blocking and ringing) and JPEG2000 (global blurring and long “waves” of distortion around contrast differences at high CR values) lead to significant changes in the statistical distribution of spectral characteristics. Furthermore, JPEG compression significantly reduces the accuracy of color resolution as a result of chroma subsampling (Shoda et al., 2023). In most cases, the strict YCbCr 4:2:0 color compression scheme is used, meaning that color resolution is halved in both rows and columns of the image, affecting the distribution of the R, G, and B color components. JPEG2000 compression mathematics also introduces distortions into the chromatic components (Marsetic et al., 2012). Unlike JPEG and JPEG2000, compression errors in AVIF and WebP do not create additional high-frequency «noise» edges (as in JPEG) or blur the image structure (as in JPEG2000). Compression artifacts are distributed along the intensity gradient, preserving the mathematical monotonicity of brightness variations. At the same time, the statistical distribution of elements within homogeneous zones changes minimally (Chlubna and Zemčík, 2025; Barman and Martini, 2020).

To study in detail the dynamics of the reduction in informativeness of spectral features, a step-by-step compression study was conducted using the WebP codec with a quality parameter Q ranging from 90 to 30 (Table 5).

Table 5. Dynamics of the generalized divergence J_x after compression of the test image S4 using WebP

Spectral feature	Uncompressed image	Q = 90	Q = 70	Q = 50	Q = 30
Channel R	14.728	9.722	10.587	12.236	14.507
Channel G	6.901	6.712	7.123	6.827	7.447
Channel B	10.448	7.423	8.176	8.113	8.438

A specific feature of the dynamics presented in Table 5 is the sharp initial drop in the feature informativeness value, even at the minimum compression ratio (Q = 90). For feature B, the generalized divergence value decreases sharply from 10.448 to 7.423, and for R – from 14.728 to 9.722. However, a further increase in the compression coefficient (a decrease in the quality parameter Q) is accompanied by a small non-monotonic increase in the divergence values, in particular, an increase in the J_x criterion to 14.507 for the R feature and to 8.438 for the B feature.

It is worth noting that this behavior of the J_x(Q) dependence is typical for almost all studied codecs, with the exception of JPEG2000 (see Fig. 6).

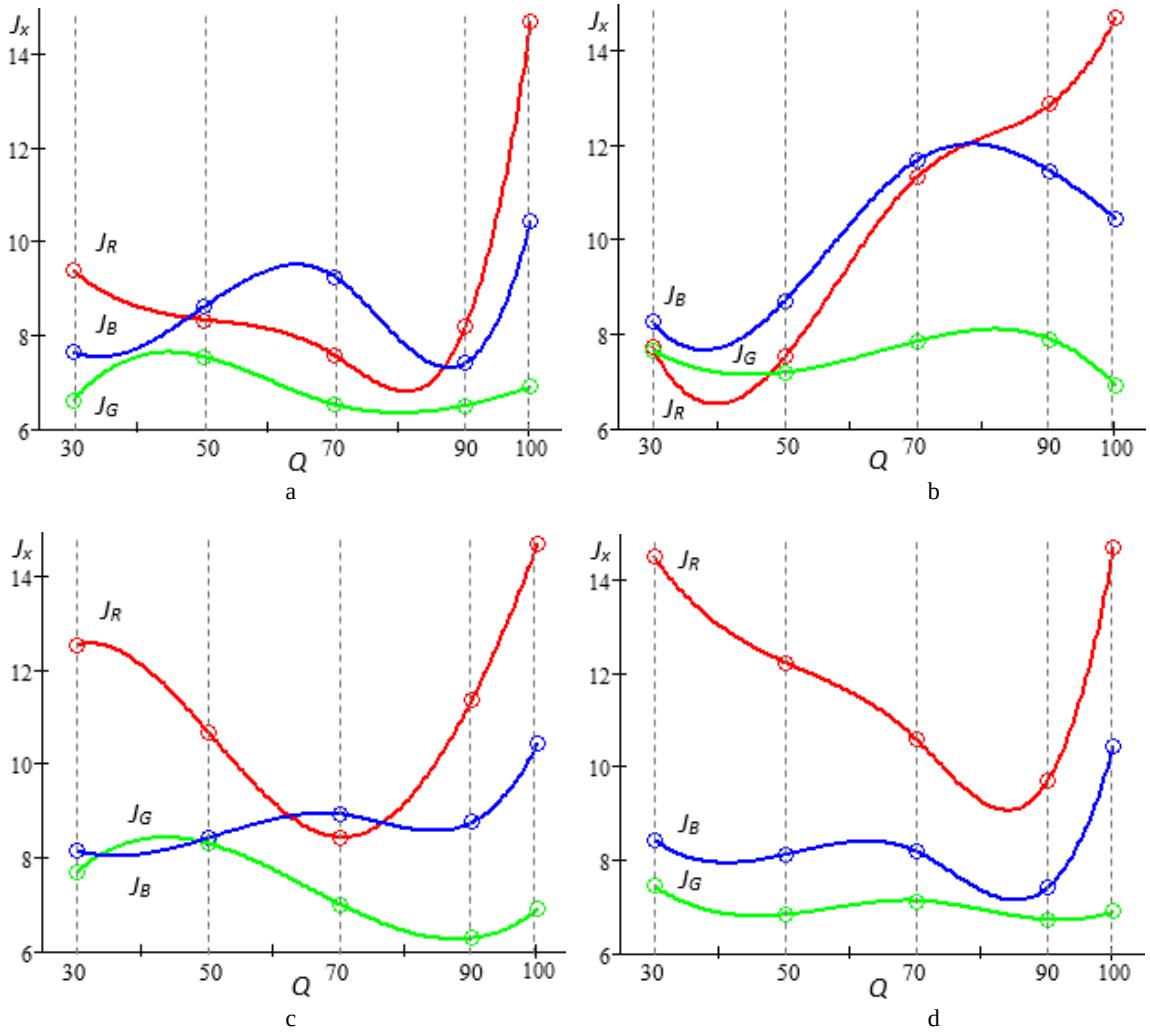


Fig. 6. Dependences of the generalized divergences J_R, J_G and J_B on the quality parameter Q for the codecs under consideration: a – JPEG, b – JPEG2000, c – AVIF, d – WebP

Spline interpolation was used to plot the J_x(Q) dependence graphs (Fig. 6); markers on the graphs show the calculated J_x criterion values for Q = 30, 50, 70, 90, 100 (Q = 100 corresponds to an uncompressed image).

To explain the nature of such nonlinear behavior of the studied informativeness criteria (Fig. 6), it is necessary to analyze the changes in the statistical characteristics of spectral features because of image compression. As an example, let us consider the case of WebP (Fig. 7). With weak compression (at Q = 90), distortions primarily lead to a significant shift in the mode of the empirical distribution and a systematic decrease in the mathematical expectation of the reference feature model (3) by approximately 4-5% for each class of objects. This leads to a geometric convergence of the scattering centers of different classes in the feature space, which in effect means a decrease in between-class distances and directly causes a decrease in divergence. On the contrary, with a further decrease in the compression quality parameter Q, the position of the distribution modes practically stabilizes, but the scale (range) of random variables begins to change noticeably. At that, the standard deviation can either increase (+35%) or decrease (–10%), which affects the distance within the class.

Therefore, if the compression algorithm reduces the variability of class features (i.e. reduces within-class scatter), the clusters of class samples become more compact, their mutual overlap decreases, which leads to an increase in their distinguishability (divergence).

Regarding the atypical behavior of the J_x(Q) dependence for features B and G in the case of JPEG2000 (see Fig. 6, b), it should first of all be noted that the anomalous increase in the generalized divergence criterion under compression with Q = 90 is associated exclusively with the increase in the local divergence of the Vegetation class (a_3). In particular, for the uncompressed image $J(3)$ = 16.21 (for B) and

$J(3) = 16.193$ (for G); with JPEG2000 compression at Q = 90, the local divergence increases to 20.331 and 19.679 for B and G, respectively. The PDFs plots (Fig. 8) for feature B illustrate the reason for this change. Compression at Q = 90 leads to a shift in the distribution for the class Vegetation to the left, into the region of lower values of B (Fig. 8, a), while for the nearest neighboring class Water (see Fig. 5, a) the distribution shifts in the opposite direction (Fig. 8, b), and, accordingly, the distinguishability of these classes increases. As the compression ratio increases, the distribution mode of feature B for the Vegetation class shifts significantly toward higher values: from 6.8 at Q = 90 to 10.4 at Q = 30. At the same time, the distribution mode for the Water class remains almost unchanged. As a result, the distance between these classes decreases and they become less distinguishable: the local divergence for feature B for the Vegetation class decreases to 12.335, and the generalized divergence decreases to 8.255 (compared to 10.448 for the uncompressed image).

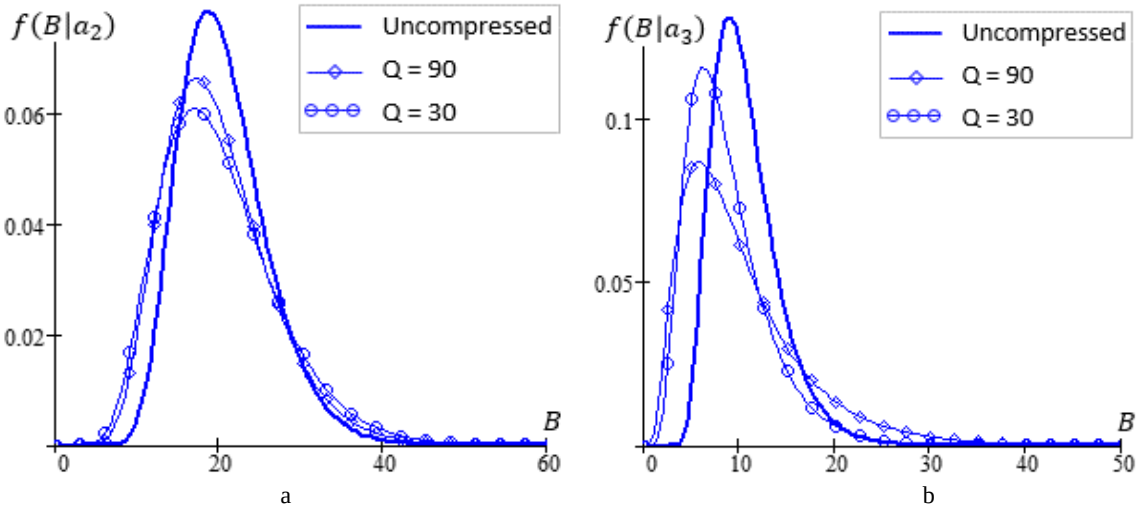


Fig. 7. Probability distributions of the spectral feature B for classes Water (a) and Vegetation (b) on uncompressed and WebP-compressed images

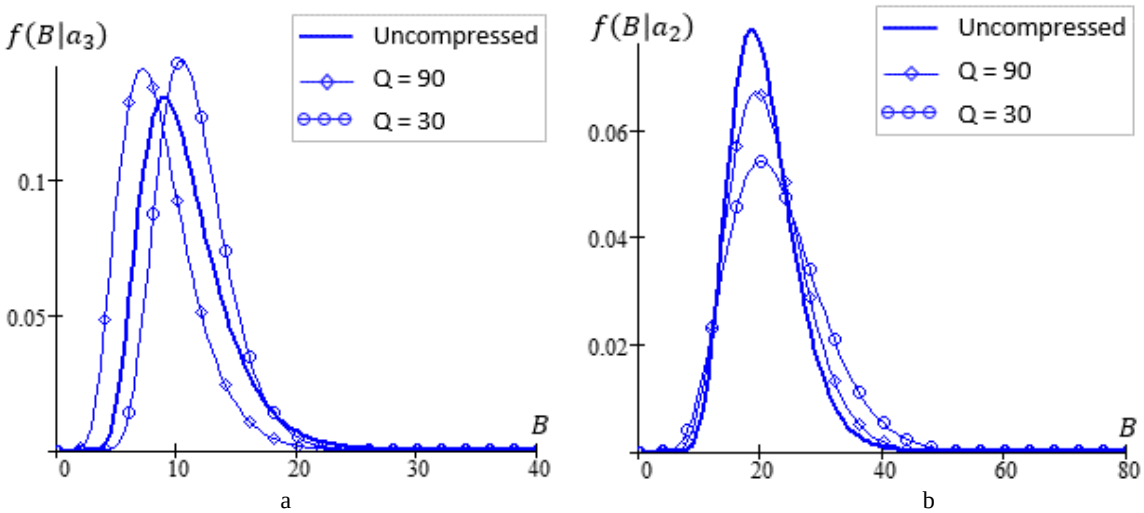


Fig. 8. Probability distributions of the spectral feature B for classes Vegetation (a) and Water (b) on uncompressed and JPEG2000-compressed images

Thus, increasing distances between classes leads to increasing divergence, whereas increasing distances within classes causes a decrease in divergence. The simultaneous presence of these two opposing trends explains the nonlinear relationship between divergence and the quality parameter Q associated with the compression ratio for the codecs under consideration.

Conclusions

The conducted research allows us to conclude that the processes of compression of multichannel remote sensing images by lossy algorithms cause a complex nonlinear deformation effect on the distributions of spectral features. The observed sharp drop in the informativeness of spectral channels at a high value of the compression quality parameter is due to the shift of statistical moments and the decrease in interclass distances in the feature space. At the same time, the effect of increasing divergence at higher compression is a mathematical artifact caused by filtering of inhomogeneities in the image due to compression, which causes a decrease in the intraclass scatter of features. Therefore, when designing automatic classification systems based on remote sensing data, the use of reference models built on uncompressed data may be

incorrect for processing compressed images, since the shift of statistical characteristics can lead to a significant decrease in recognition accuracy even at relatively small compression ratios.

Author Contributions: Conceptualization – I. K. Vasylieva and S. Y. Shulhin; Methodology – S. Y. Shulhin and I. K. Vasylieva; Formal Analysis and Data Processing – S. Y. Shulhin; Investigation – S. Y. Shulhin; Writing – Original Draft Preparation – S. Y. Shulhin and I. K. Vasylieva; Writing – Review & Editing – S. Y. Shulhin and I. K.Vasylieva. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Disclosure of AI use: We confirm that no generative artificial intelligence tools were used in the preparation of this manuscript.

Data Availability Statement: Data available on reasonable request from the authors.

Acknowledgments: The authors are grateful to the National Academy of Sciences of Ukraine for supporting this research. We are also grateful to the reviewers and editors for their valuable comments, recommendations, and attention to the work.

Conflicts of Interest: The authors declare no conflict of interest

Внесок авторів: Концептуалізація – І. К. Васильєва та С. Ю. Шульгін; Методологія – С. Ю. Шульгін та І. К. Васильєва; Формальний аналіз і обробка даних – С. Ю. Шульгін; Розслідування – С. Й. Шульгін; Написання – Підготовка оригіналу – С. Ю. Шульгін та І. К. Васильєва; Написання – рецензування та редагування – С. Ю. Шульгін та І. К. Васильєва. Усі автори прочитали та погодилися з опублікованою версією рукопису.

Фінансування: це дослідження не отримувало зовнішнього фінансування.

Розкриття використання штучного інтелекту: ми підтверджуємо, що під час підготовки цього рукопису не використовувалися інструменти генеративного штучного інтелекту.

Доступність та цілісність даних: дані доступні за обгрунтованим запитом від авторів.

Подяки: Автори вдячні Національній академії наук України за підтримку цього дослідження. Ми також вдячні рецензентам і редакторам за цінні зауваження, рекомендації та увагу до роботи.

Конфлікт інтересів: автори заявляють про відсутність конфлікту інтересів

References

Abramova, V., Lukin, V., Abramov, S., Abramov, K., & Bataeva, E. (2022). Analysis of Statistical and Spatial Spectral Characteristics of Distortions in Lossy Image Compression. In *2022 IEEE 2nd Ukrainian Microwave Week (UkrMW)*, 644–649. IEEE. doi:10.1109/ukrmw58013.2022.10036949.

Barina, D. (2021). Comparison of Lossless Image Formats. *Proceedings of 29th Int. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision 2021*, 1–4. doi:10.48550/arXiv.2108.02557.

Barman, N., Martini, M. G. (2020). An Evaluation of the Next-Generation Image Coding Standard AVIF. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. doi:10.1109/qomex48832.2020.9123131.

Chlubna, T., Zemčík, P. (2025). Comparative survey of image compression methods across different pixel formats and bit depths. *Signal, Image and Video Processing*, 19(12). doi:10.1007/s11760-025-04579-6.

Corino, M. G., Leidi, T., & Peternier, A. (2026). AVIF as an Alternative to JPEG and GPU Texture Compression Schemes for Texture Storage in 3D Computer Graphics. *Applied Sciences*, 16(5), 2541. doi:10.3390/app16052541.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). John Wiley & Sons.

Gandor, T., Nalepa, J. (2022). First Gradually, Then Suddenly: Understanding the Impact of Image Compression on Object Detection Using Deep Learning. *Sensors*, 22(3), 1104. doi:10.3390/s22031104.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37. doi:10.1109/34.824819.

Johnson, N. L. (1949). Systems of Frequency Curves Generated by Methods of Translation. *Biometrika*, 36(1–2), 149–176. doi:10.1093/biomet/36.1-2.149.

Kovalenko, B., Lukin, V. (2023). Analysis of distortions due to BPG-based lossy compression of noise-free and noisy images. *Herald of Khmelnytskyi National University. Technical sciences*, 325(5), 128–135. doi:10.31891/2307-5732-2023-325-5-128-135.

Kullback, S., Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. http://www.jstor.org/stable/2236703.

Lukin, V., Vasilyeva, I., Krivenko, S., Li, F., Abramov, S., Rubel, O., Vozel, B., Chehdi, K., & Egiazarian, K. (2020). Lossy Compression of Multichannel Remote Sensing Images with Quality Control. *Remote Sensing*, 12(22), 3840. doi:10.3390/rs12223840.

Lukin, V., Vasilyeva, I., & Vozel, B. (2022). Results of Approbation of the Method for Predicting the Classification Accuracy of Multichannel Images. In *2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC)*. IEEE. doi:10.1109/saic57818.2022.9923028.

Marsetic, A., Kokalj, Z., & Ostir, K. (2012). The Effect of Lossy Image Compression on Object Based Image Classification – Worldview-2 Case Study. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-4/W19, 187–192. doi:10.5194/isprsarchives-XXXVIII-4-W19-187-2011.

Pomarico, E., Schmidt, C., Chays, F., Nguyen, D., Planchette, A., Tissot, A. … Extermann, J. (2022). Statistical distortion of supervised learning predictions in optical microscopy induced by image compression. *Scientific Reports*, 12(1), 3464. doi:10.1038/s41598-022-07445-4.

Shoda, A., Miyazaki, T., & Omachi, S. (2023). JPEG Image Enhancement with Pre-Processing of Color Reduction and Smoothing. *Sensors*, 23(21), 8861. doi:10.3390/s23218861.

Shulhin, S., Vasylieva, I. (2025). Statistical Analysis of the Impact of Lossy Compression Algorithms on the Structure of Digital Images. *Herald of Khmelnytskyi National University. Technical sciences*, 359(6.2), 22–30. doi:10.31891/2307-5732-2025-359-73.

Sultana, Z., Nahar, L., Tasnim, F., Hossain, M. S., & Andersson, K. (2022). Lossy Compression Effect on Color and Texture Based Image Retrieval Performance. In *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, 569, 1159–1167. Springer, Cham. doi:10.1007/978-3-031-19958-5_108.

Vasil'eva, I., Popov, A. (2016). An algorithm for recognition of hydrometeors by polarimetric radar data based on the information theory. In *2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW)*. IEEE. doi:10.1109/msmw.2016.7538144.

АНАЛІЗ ВПЛИВУ СТИСНЕННЯ БАГАТОКАНАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ З ВТРАТАМИ НА ІНФОРМАТИВНІСТЬ СПЕКТРАЛЬНИХ ОЗНАК

Шульгін С. Ю., https://orcid.org/0009-0005-3997-3998
Васильєва І. К., http://orcid.org/0000-0002-1378-1104
Кафедра інформаційно-комунікаційних технологій, Національний аерокосмічний університет, 61070 Харків, Україна

Розвиток технологій дистанційного зондування та висока просторова роздільна здатність сучасних супутникових систем призвели до збільшення обсягів даних спостережень. Проблема зберігання та передавання великих обсягів даних зумовлює необхідність використання методів стиснення. Однак стиснення з втратами неминуче призводить до спотворень статистичної структури спектральних компонент даних, що впливає на результати їх подальшого аналізу та інтерпретації. У статті викладено результати дослідження впливу найпоширеніших технологій стиснення (JPEG, JPEG2000, AVIF, WebP) на інформативність спектральних ознак для статистичної класифікації типів земної поверхні на основі даних Sentinel-2. Для опису емпіричних розподілів ознак використано S_B-розподіл Джонсона, що враховує негауссівський характер розподілів та високий рівень крос-кореляції між каналами R, G і B. Кількісне оцінювання інформативності ознак базується на критеріях локальної та узагальненої дивергенцій. Аналіз зображень з різною структурною складністю виявив нелінійний характер спотворень статистичних моментів еталонних класів під впливом стиснення з втратами. Для зображень з відносно простою структурою алгоритми кодування зміщують моду розподілів, зменшуючи розсіювання всередині класу, що може призвести до збільшення узагальненої дивергенції. І навпаки, для зображень з високою просторовою деталізацією навіть найменше стиснення призводить до значного зниження інформативності ознак внаслідок зменшення міжкласових відстаней. Ці результати підкреслюють необхідність розроблення адаптивних класифікаторів для стиснутих зображень, які можуть динамічно компенсувати статистичні спотворення, специфічні для кодеків.

Ключові слова: стиснення з втратами, багатоканальні космічні зображення, інформативність, узагальнена дивергенція, S_B-розподіл Джонсона, статистичний аналіз.

Рукопис статті отримано 16.08.2026
Надходження остаточної версії: 11.09.2026
Публікація статті: 30.09.2026
Received 16.08.2026
Revised 11.09.2026
Accepted 30.09.2026