An Information Model for Digital Image Segmentation

D. M. Murashov*

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44/2, Vavilov Street, Moscow, 119333 Russia

**e-mail: d_murashov@mail.ru*

Abstract—This paper investigates an iterative information-theoretical method for segmentation of digital images. A system that includes a segmentation algorithm with a parameter that determines the number of image segments and a procedure for setting the value of this parameter that minimizes the information redundancy measure is considered. A new simplified mathematical model is proposed to analyze the properties of this system. It is shown that there exists a minimum of the redundancy measure for the proposed model. The adequacy of the model is confirmed experimentally. The computational experiment carried out on images from the Berkeley Segmentation Dataset (BSDS500) shows that a segmented image corresponding to the minimum redundancy measure has the highest informational similarity to ground truth segmentations available in BSDS500. We compared the image segmentation results provided by the EDISON system using the minimum information redundancy criterion and entropy criterion. The advantage of the minimum redundancy criterion is demonstrated.

Keywords: image segmentation, segmentation quality, information redundancy measure, variation of information

DOI: 10.1134/S1054661821040179

INTRODUCTION

Image segmentation consists in partitioning a digital image, represented as a region, into a set of non-overlapping connected subregions (segments) the elements of which are similar with respect to some characteristic and differ from the elements of adjacent regions [14, 15]. In this case, the problem of choosing parameters of image segmentation algorithms arises. These parameters are set to obtain the best image partition. The quality of the partition can be evaluated by an expert (visual estimation) or based on a quantitative index.

Survey [33] provides a rather complete classification of methods for evaluating image segmentation quality. Supervised and unsupervised methods can be distinguished as the main groups of these methods. A separate group includes methods that use classifiers trained on certain sets of features extracted from input images; these classifiers are capable of predicting image segmentation results.

With supervised methods, the segmentation result is usually compared with an image segmented by a specialist and taken as a ground truth segmentation [2]. There can be several ground truth segmentations provided by different experts. The quality of image segmentation can be characterized by various measures that describe edge detection error, consistency of segments, overlap of segments (Szymkiewicz–Simpson coefficient), etc. In [2, 20], the precision and recall measures were used to compare segment boundaries. In [19], Martin et al. proposed global and local consistency errors as measures for comparing segments in the output and ground truth segmentations. Other measures for evaluating segmentation quality were discussed in [6, 33]. If the image segmentation operation is considered as a process of pixel clustering, then set-theoretical, statistical, and information-theoretical measures [33] are used to compare data clustering results. The most popular measures are the Chi-square, the Rand index [26] and its modifications [31], the Jaccard index [33], the Fowlkes–Mallows measure [10], mutual information and normalized mutual information [30], and variation of information [21]. These measures allow one to compare different variants of image segmentation.

Based on similarity/dissimilarity measures, a number of automated image segmentation methods were developed. In [7, 18], empirical quality functionals were proposed that take into account the number of image segments and variation of color characteristics within the segments of output images.

In [27], a two-step algorithm for segmentation of MR and CT images was proposed; it represents the image segmentation process by using models of direct and reversed information channels. At the first step, the image is structured into quasi-homogeneous regions, based on the condition of maximum mutual information between the input image and the resulting partition. At the second step, the intensity levels of the

Received June 17, 2021; revised June 17, 2021; accepted June 17, 2021

input image histogram are clustered based on the minimum loss of mutual information between the clustering result and the partition obtained at the first step. Since the dependences of the mutual information at the input and output of the direct and reversed channels do not have extreme properties, the authors used additional conditions to obtain the best segmentation result (the number of segments, the probability of error, and the ratio between the mutual information values of the reversed and direct channels).

In [32], an information-theoretic method for evaluation of image segmentation results was developed. The entropy-based quality measure takes into account the heterogeneity of pixel characteristics within segments and the complexity of partitioning the image into segments. According to the authors, the best image segmentation results correspond to the local minima of the entropy-based measure. This measure allows one to compare the qualities of different segmentations obtained with the same algorithm at different parameter values, as well as compare the results provided by different algorithms.

A heuristic method based on the iterative application of the mean shift algorithm [8] was described in [29]. As a criterion for selecting the best partition, the relative rate of decrease in the entropy of the image obtained at the filtering stage between iterations of the mean shift algorithm with different parameter values is used. The threshold value for this criterion is chosen empirically. A disadvantage of this method is the lack of universality as it is oriented to a particular image segmentation algorithm.

An alternative approach is based on fusion of segmented images, rather than on finding an image partition that optimizes a quality measure selected. In [22], an iterative method was proposed that fuses a set of coarsely segmented images obtained in different color spaces for different values of parameters of a clustering algorithm. This idea was successfully implemented to fuse ensembles of data partitions obtained by clustering [12, 30]. For evaluation of image segmentation quality, the criterion that characterizes the mean variation of information between a fused image and each coarse segmentation was used. In [16], a generalization of the method [22] was proposed. The problem of fusing coarsely segmented images is solved as a multicriterion optimization problem. To evaluate the result of image segment fusion, the minimum mean variation of information is computed; to evaluate the accuracy of segment boundaries, the F-measure is used. The multicriterion approach improves segmentation of images from the Berkeley Segmentation Dataset (BSDS500) as compared to the single-criteria method.

The group of methods considered above consists mainly of heuristic methods, which provide good results on test sets of images.

A number of publications were devoted to automatic image segmentation based on classifiers, which are used to predict quality of image segmentation. In [17], a method for segmentation error evaluation based on a regression algorithm was proposed. The method consists of the following stages: computing the features that characterize the image segmentation result and training the regression algorithm on a set of images with available ground truth segmentations to predict the error. The method can be used to set the parameters of the image segmentation algorithm or select the best result when analyzing the input image by several segmentation algorithms running in parallel. In [13], when setting the parameters of the algorithm, the similarity between the segmentation result and the original image was estimated. As a similarity measure, the weighted uncertainty index was used; it is computed based on values of normalized mutual information for the corresponding color channels of the input and segmented images [11]. Based on expert estimates of segmentation results for the series of images, obtained at different parameter values on the coordinate plane by using the classifier, regions of undersegmentation, oversegmentation, and optimal segmentation were identified. In the process of image segmentation, the parameter of the algorithm at each point of the processed image is set using an iterative procedure based on the graph-cut algorithm [9]. In recent years, methods for predicting the quality of image segmentation based on deep neural networks have appeared [28]. In this case, the Dice similarity coefficient is used as a measure.

The disadvantages of the classifier-based approach are the subjectivity of expert estimates and the fact that the trained algorithm provides acceptable results only for those classes of images that were used in the process of training.

In [4], an information-theoretical model of the human visual system was proposed. The model is based on Barlow's hypothesis [5] about minimization of information redundancy at the initial stages of signal processing in the human visual system. It was assumed that, at the initial stages of visual perception, the redundancy of the information coming from the retina through the optic nerve is reduced. In [24, 25], using the principle of redundancy minimization [4, 5], the criterion based on the minimum of the information redundancy measure was used to find the best image partition in a set of partitions obtained at different values of the segmentation parameter. This paper investigates this criterion of image segmentation quality. To analyze the properties of the criterion, a simplified mathematical model of the image segmentation process is proposed. It is shown that there exists a minimum of the redundancy measure for the proposed model. To confirm the validity of the proposed model, a computational experiment is carried out on images from BSDS500 [2]. The results of image segmentation based on the criterion of minimum redundancy and the entropy criterion described in [29] are compared.

1. MODEL OF THE IMAGE SEGMENTATION SYSTEM

To investigate the properties of the image segmentation system, it is required to construct its mathematical model. In this section, we propose and analyze a simplified information model that is not oriented to any particular segmentation algorithm.

1.1. Image Segmentation Problem

The image segmentation operation can be represented by the following model [24]:

$$V = F(U, t), \tag{1}$$

where $U : \mathbb{Z}^2 \to \mathbb{Z}$ is the input image, $V : \mathbb{Z}^2 \to \mathbb{Z}$ is the segmented image, F is the operator that describes the segmentation algorithm, and $t \in \mathbb{R}$ is the segmentation parameter. The image segmentation problem can be formulated as follows. Given input image U, segmentation algorithm (1) yields a set Q of images $\mathfrak{V} = \{V_1, V_2, \ldots, V_q, \ldots, V_Q\}$ for different values of parameter t. It is required to select image $V_{q\min}$ that minimizes measure $M(U, V_q)$:

$$q_{\min} = \arg\min_q \left[M\left(U, V_q\right)\right], \quad q = 1, 2, \ldots, Q. \tag{2}$$

In the next section, we use the information criterion of image segmentation quality proposed in [24, 25].

1.2. Information Criterion of Image Segmentation Quality

In [4], the criterion of minimum information redundancy was used as a basis for the information-theoretical model of the human visual system. In [24], the criterion of minimum information redundancy was implemented in an image segmentation algorithm. To apply the information-theoretical approach, a probabilistic model of the relationship between the original and segmented images is required.

For simplicity, we consider the process of grayscale image segmentation. Suppose that the original and segmented images are the input and output of the stochastic information system. The gray levels of the images are described by discrete random variables U and V with values u and v. Variable U has L gray levels, while V can have $1 \le l \le L$ gray levels.

The image segmentation operation is represented by the following information channel model, which is similar to the model used in [23]:

$$V = F(U + \eta), \tag{3}$$

where U is the signal at the input of the channel, V is the output of the channel, F is a transformation function, and η is channel noise. Variables V and η are assumed to be independent.

We use the redundancy measure [4] as a measure of image segmentation quality:

$$R = 1 - \frac{I(U;V)}{C(V)} \tag{4}$$

where $I(U;V)$ is mutual information and $C(V)$ is channel capacity. Suppose that $C(V) = H(V)$, where $H(V)$ is the entropy of the output. Then, taking into account that $I(U;V) = H(V) - H(V|U)$, expression (4) takes the following form:

$$R = \frac{H(V|U)}{H(V)}, \tag{5}$$

where $H(V|U)$ is the conditional entropy of the channel output given input U. The criterion of image segmentation quality is the minimum channel redundancy.

In the following section, we propose a simplified qualitative model of grayscale image segmentation and show that this model provides the minimum of the redundancy measure.

1.3. Information Model of Image Segmentation

To investigate the qualitative properties of the redundancy measure, we need to construct a model of joint two-dimensional discrete distribution of gray levels for the input and output of the segmentation system. This model allows us to analyze the dynamics of information measures (entropies) that characterize the process of image segmentation.

To construct the qualitative model, it is required to consider the dynamics of the joint two-dimensional discrete distribution of gray levels for the input and output of the segmentation system depending on the number of segments K in image V. Suppose that grayscale image U shown in Fig. 1a is input to the image segmentation algorithm. Figures 1b–1d show two-dimensional discrete distributions for three values of K and number of grayscale levels $L = 16$ in image U. The figures illustrate the distributions of gray tones in the image segments. Each segment has a significant number of pixels of the dominant gray level, which form the peaks of the distribution, and some pixels with other gray levels. When the number of image segments is large, the distribution of gray levels within segments has sharp peaks. With enlargement of the segments and reduction in their number, the peaks are smoothed out (see Fig. 1).

Suppose that the joint distribution of gray levels for images U and V can be represented by K components corresponding to segments of image V. For simplicity,

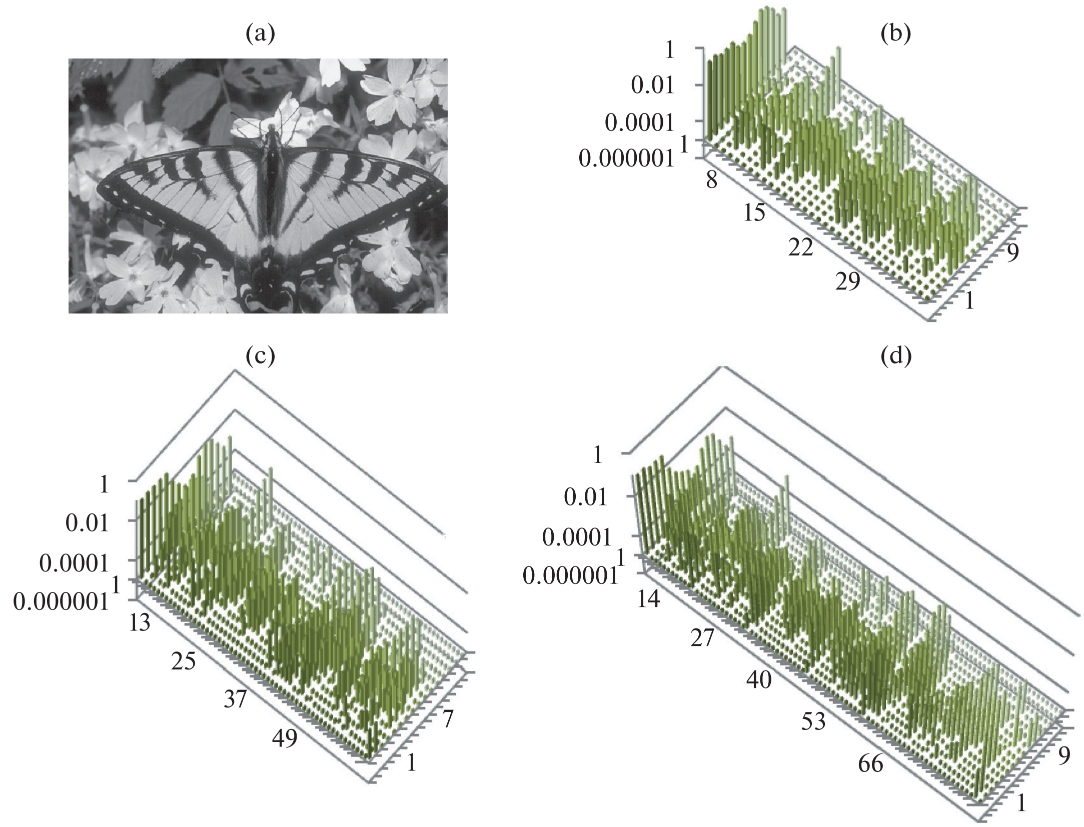

Fig. 1. Joint discrete distribution of gray levels for the input and output of the image segmentation system at three different values of K (number of image segments): (a) grayscale version of the input image (35010.jpg). Joint discrete distribution of gray levels for U and V at $K =$ (b) 32, (c) 57, and (d) 75.

we assume that all components are of the same type and consist of L elements that correspond to the frequency of occurrence of pixels with gray level l in the kth segment of V. Each of the components has a peak corresponding to the dominant gray level.

Suppose that the components have different dominant gray levels. Suppose also that $P(u_l, v_k) = P$ if $l = k$. The relationship between the probabilities of gray levels l in the components is determined by coefficient α, $0 < \alpha \le 1$. For instance, in the segment of V encoded by level v_1, we have

$$\begin{aligned} P(u_2, v_1) &= P(u_3, v_1) = \ldots = P(u_L, v_1) \\ &= \alpha P(u_1, v_1) = \alpha P, \end{aligned} \tag{6}$$

where $\alpha = \alpha(K)$ depends on the number of segments in image V. The model of the two-dimensional discrete distribution described above is shown in Fig. 2. For this model, the following relations hold: $P(v_k) = (L-1)\alpha(K)P + P$ and $K\left[(L-1)\alpha(K)+1\right]P = 1$, which imply that $P = 1/K\left[(L-1)\alpha+1\right]$. Then, taking into account model (6) of the joint discrete distribution of gray levels for the input and output of the image segmentation system, we express the entropies included in formula (5) as follows:

$$\begin{aligned} H(U,V) &= -\sum_{l=1}^{L}\sum_{k=1}^{K} P(u_l, v_k)\log P(u_l, v_k) \\ &= \log K + \log\left[\alpha(L-1)+1\right] - \frac{(L-1)\alpha\log\alpha}{\alpha(L-1)+1} \end{aligned} \tag{7}$$

$$\begin{aligned} H(U) &= -\sum_{l=1}^{L}\left\{\left[\sum_{k=1}^{K} P(u_l, v_k)\right]\log\left[\sum_{k=1}^{K} P(u_l, v_k)\right]\right\} \\ &= -LP_0\left[(K_0-1)\alpha_0+1\right]\log\left[(K_0-1)\alpha_0 P_0 + P_0\right]; \end{aligned} \tag{8}$$

$$\begin{aligned} H(V) &= -\sum_{k=1}^{K}\left\{\left[\sum_{l=1}^{L} P(u_l, v_k)\right]\log\left[\sum_{l=1}^{L} P(u_l, v_k)\right]\right\} \\ &= \log K, \end{aligned} \tag{9}$$

$$\begin{aligned} H(V|U) &= H(U,V) - H(U) \\ &= \log K + \log\left[\alpha(L-1)+1\right] - \frac{(L-1)\alpha\log\alpha}{\alpha(L-1)+1} \\ &+ LP_0\left[(K_0-1)\alpha_0+1\right]\log\left[(K_0-1)\alpha_0 P_0 + P_0\right], \end{aligned} \tag{10}$$

where P_0, K_0, and α_0 are quantities that correspond to image U.

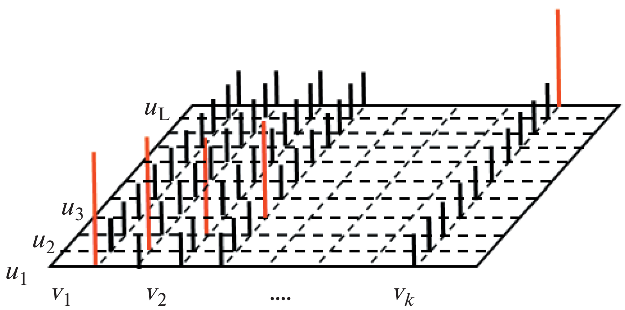

Fig. 2. Model of the joint discrete distribution of gray levels for the input and output images (U and V).

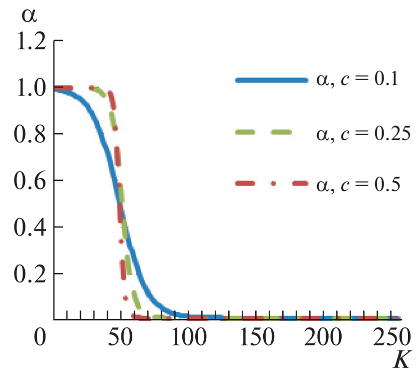

Fig. 3. Function $\alpha(K)$ at different values of parameter c.

By substituting expressions (7)–(10) into (5), we obtain

$$R = 1 + \frac{\log[\alpha(L-1)+1] - \dfrac{(L-1)\alpha\log\alpha}{\alpha(L-1)+1} + LP_0[(K_0-1)\alpha_0+1]\log[(K_0-1)\alpha_0 P_0 + P_0]}{\log K}. \quad (11)$$

It should be noted that, with decreasing number of segments K in image V, the distribution (see Fig. 1) changes. In the model shown in Fig. 2, this transformation of the distribution corresponds to an increase in coefficient α. This dependence can be represented by a monotonic function:

$$\alpha(K) = \frac{a}{1+e^{c(K-b)}} + d, \quad (12)$$

where a, b, c, and d are parameters ($b > 1$ and $d = 1 - a$). The behavior of dependence $\alpha(K)$ for different values of parameter c is shown in Fig. 3.

Let us show that redundancy measure R defined by expressions (11), (7)–(10), and (12) has a minimum. Variable K takes values from a set of integers. To analyze function R, in formulas (7)–(12) we replace integer variable K with real variable z. For simplicity, we assume that $\log(x)$ in expression (11) is a natural logarithm. Then, derivative dR/dz has the following form:

$$R'(z) = -\frac{\log[\alpha(L-1)+1]}{z\log^2 z} + \frac{\alpha(L-1)\log\alpha}{z[\alpha(L-1)+1]\log^2 z} + \frac{ace^{c(z-b)}(L-1)\log\alpha}{[\alpha(L-1)+1]^2[e^{c(z-b)}+1]^2\log z} + \frac{H(U)}{z\log^2 z}. \quad (13)$$

Let us introduce the following designations:

$$R_1'(z) = -\frac{\log[\alpha(L-1)+1]}{z\log^2 z};$$

$$R_2'(z) = \frac{\alpha(L-1)\log\alpha}{z[\alpha(L-1)+1]\log^2 z};$$

$$R_3'(z) = \frac{ace^{c(z-b)}(L-1)\log\alpha}{[\alpha(L-1)+1]^2[e^{c(z-b)}+1]^2\log z}; \quad (14)$$

$$R_4'(z) = \frac{H(U)}{z\log^2 z}.$$

We consider the behavior of function $R'(z)$ when changing z and, therefore, $\alpha(z)$. Suppose that $K_0 = L$; then, $H(U) = \log(K_0) = \log(L)$. For $z = K_0$, $\alpha_0 = \alpha(K_0) = d$, $d \ll a$. Function $\alpha(z)$ described by expression (12) decreases with increasing z (see Fig. 3).

For small z and $\alpha \approx 1$, the derivative of the redundancy is close to zero, $R'(z) \approx 0$.

Suppose that $z = b$ and $\alpha(b) = \alpha(0)/2 = 0.5$. Expression (13) implies

$$R'(z) = \frac{(\alpha L - \alpha + 1)\log\dfrac{L}{\alpha L - \alpha + 1} + \alpha(L-1)\log\alpha}{(\alpha L - \alpha + 1) z \log^2 z} + \frac{ace^{c(z-b)}(L-1)\log\alpha}{[\alpha(L-1)+1]^2 [e^{c(z-b)}+1]^2 \log z}$$
$$= \frac{4[\alpha(L-1)+1]^2 \log\dfrac{L}{\alpha L - \alpha + 1} + 4[\alpha^2(L-1)^2 + \alpha(L-1)]\log\alpha + acb(L-1)\log b\log\alpha}{4[\alpha(L-1)+1]^2 b\log^2 b}$$
$$= \frac{(L+1)^2\log\dfrac{2L}{L+1} - (L^2-1)\log 2 - acb(L-1)\log b\log 2}{(L+1)^2 b\log^2 b}.$$

Inequality $R'(z) < 0$ holds if

$$(L+1)^2\log\frac{2L}{L+1} - (L^2-1)\log 2 - acb(L-1)\log b\log 2 < 0,$$

or

$$acb(L-1)\log b\log 2 > (L+1)^2\log\frac{2L}{L+1} - (L^2-1)\log 2. \tag{15}$$

Let us strengthen condition (15):

$$acb(L-1)\log b\log 2 > (L+1)^2\log 2 - (L^2-1)\log 2.$$

In this case, inequality $R'(z) < 0$ holds if

$$acb\log b > 2\frac{L+1}{L-1}. \tag{16}$$

For instance, at $L = 256$, $a = 0.99$, and $c = 0.1$, condition (16) and, therefore, condition (15) hold if $b \geq 10$; at $c = 0.5$, condition (16) holds if $b \geq 6$.

Suppose that $z = K_0 = L$ and $\alpha = d$. Assume also that $d = 1/L$. Let us find conditions under which $R'(z) > 0$. Based on the assumptions made above, (13) and (14) imply

$$R'(z) = \frac{(2L-1)\log\dfrac{L^2}{(2L-1)} - (L-1)\log L}{L(2L-1)\log^2 L} - \frac{ace^{c(L-b)}(L-1)L^2}{(2L-1)^2[e^{c(L-b)}+1]^2}. \tag{17}$$

For $R'(z) > 0$ to hold, the following inequality must hold:

$$\frac{(2L-1)\log\dfrac{L^2}{(2L-1)} - (L-1)\log L}{L(2L-1)\log^2 L} - \frac{ace^{c(L-b)}(L-1)L^2}{(2L-1)^2[e^{c(L-b)}+1]^2} > 0 \tag{18}$$

Let us find the relationship of parameters of the model (6)–(12) in such a way that inequality (18) holds.

Let us strengthen inequality (18):

$$\frac{(2L-1)\log\dfrac{L^2}{2L-1} - (L-1)\log L}{L(2L-1)\log^2 L} - \frac{ac(L-1)L^2}{(2L-1)^2[e^{c(L-b)}+1]} > 0.$$

By transformation, we obtain

$$\frac{(2L-1)\left[(2L-1)\log\dfrac{L^2}{(2L-1)} - (L-1)\log L\right][e^{c(L-b)}+1] - ac(L-1)L^3\log^2 L}{L(2L-1)^2[e^{c(L-b)}+1]\log^2 L} > 0. \tag{19}$$

Inequality (19) holds if

$$(2L-1)\left[(2L-1)\log\frac{L^2}{(2L-1)} - (L-1)\log L\right][e^{c(L-b)}+1] - ac(L-1)L^3\log^2 L > 0,$$

which leads to the condition

$$b < L - \frac{1}{c}\log\left\{\frac{ac(L-1)L^3\log^2 L}{(2L-1)\left[(2L-1)\log\frac{L^2}{(2L-1)} - (L-1)\log L\right]} - 1\right\}. \tag{20}$$

For instance, at $L = 256$, $a = 0.99$, and $c = 0.1$, condition (18) holds if $b < 155$; at $c = 0.5$, condition (18) holds if $b < 232$.

Thus, in a fairly wide range of the parameters of model (6)–(12), the value of $R'(z)$ is negative at relatively small z and $\alpha < 1$, and it is positive at large z and $\alpha \approx d$. Therefore, there is at least one point z at which $R'(z) = 0$ with the second derivative at this point being positive, $R''(z) > 0$. This means that function $R(z)$ has a minimum.

This result can be formulated as follows.

Statement. Suppose that the image segmentation system is described by model (1)–(3), (5), (6), (11), and (12). Then, for the model parameters determined by conditions (16) and (20), information redundancy measure (5) has a minimum.

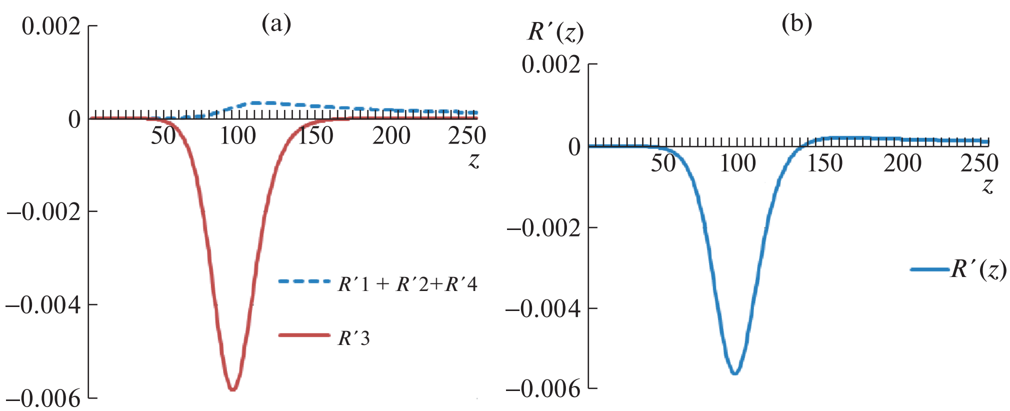

Fig. 4. Function $R'(z)$ and its components at $b = 50$, $c = 0.1$, and $L = 256$: (a) functions $R_1'(z) + R_2'(z) + R_4'(z)$ and $R_3'(z)$ and (b) function $R'(z)$.

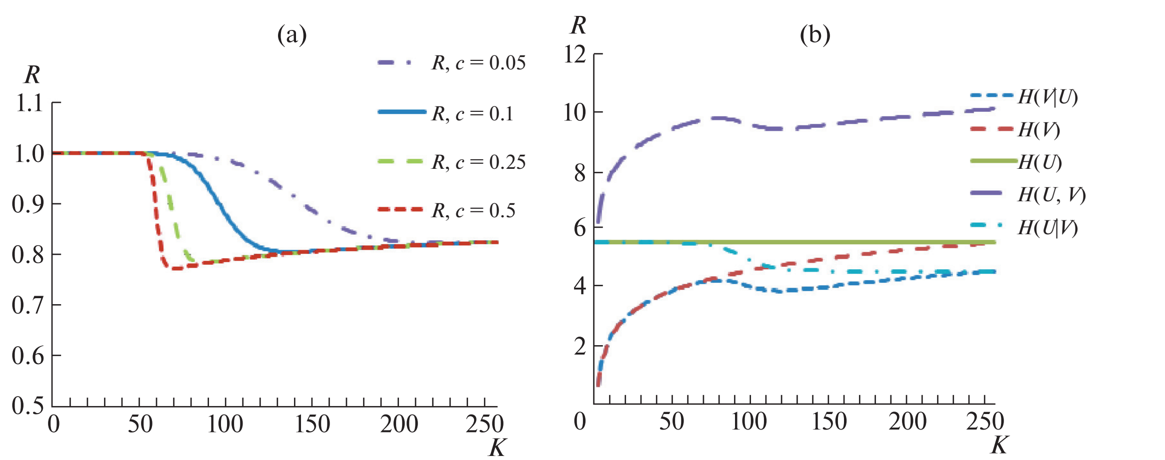

Fig. 5. Dependence of redundancy measure $R(K)$ and entropies for the input and segmented images: (a) function $R(K)$ at different values of parameter c and (b) entropies that determine the value of information redundancy R vs. number of image segments K at $c = 0.1$.

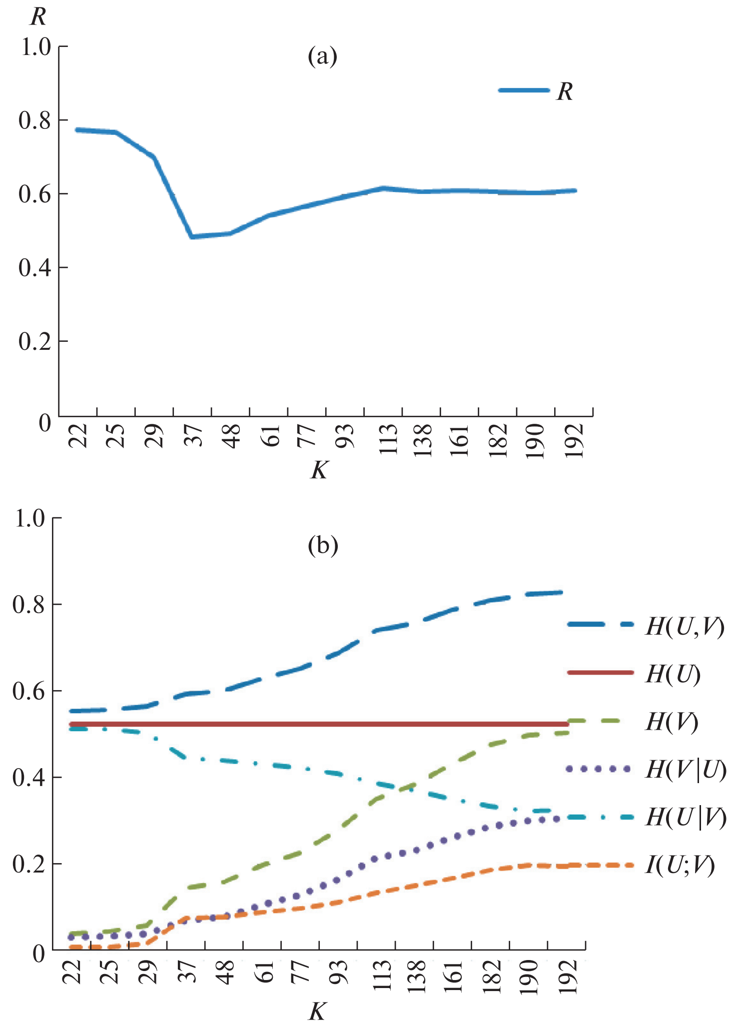

Fig. 6. Dependences of the information characteristics for the image from Fig. 1a: (a) function $R(K)$ and (b) entropies and average mutual information $I(U;V)$ that determine $R(K)$.

The graphs of functions $R_1'(z)+R_2'(z)+R_4'(z)$, $R_3'(z)$, and $R'(z)$ (see (14)) at $b = 50$, $c = 0.1$, and $L = 256$ (see Figs. 4a and 4b) confirm this statement.

The graphs of function $R(K)$, given by expression (11), at various values of parameter c for dependence $\alpha(K)$, given by expression (12), are shown in Fig. 5a. It can be seen from Fig. 5a that function $R(K)$ has a minimum. Variation of entropies (7)–(10), which determine the value of information redundancy (11), is shown in Fig. 5b versus the number of image segments.

Figures 6a and 6b show variations in the information redundancy and entropies that determine information redundancy (5) versus the number of image segments for the image depicted in Fig. 1a. Figures 4a and 6a, as well as 4b and 6b, demonstrate the qualitative similarity in the dynamics of the information redundancy and entropy values depending on the number of image segments when partitioning the hypothetical and real images.

Now, let us show that segmented image $V(K_{\min})$, which corresponds to minimum redundancy $R(K_{\min})$, has a sufficiently small informational difference from original image U.

As a measure of difference, we use variation of information, which is a metric proposed in [21]; it has the properties necessary to compare data clustering results. In our case, the variation of information characterizes the difference (distance) between the original and segmented images:

$$VI(U,V) = H(U,V) - I(U;V), \tag{21}$$

where $VI(U,V)$ is the variation of information, $H(U)$ and $H(V)$ are the entropies of the compared images

(U and V), and $I(U;V)$ is the mutual information of the compared images.

To measure the difference between the original (U) and segmented (V) images, we use the normalized variation of information:

$$VI_n(U,V) = \frac{VI(U,V)}{H(U,V)}, \tag{22}$$

where $VI_n(U,V)$ is the normalized variation of information. Let us evaluate $VI_n(U,V)$ for segmented image $V(K_{\min})$. For this purpose, using formulas (5) and (21), as well as relation $H(U,V) = H(U) + H(V|U)$, we express $VI_n(U,V)$ in terms of information redundancy $R(K)$:

$$VI_n(U,V) = \frac{VI(U,V)}{H(U,V)} = 1 - \frac{[1-R(K)]}{\dfrac{H(U)}{H(V)} + R(K)}. \tag{23}$$

As above, we replace integer variable K with real variable z and compute derivative of $VI_n(U,V)$ with respect to z:

$$\begin{aligned} VI_n'(U,V) &= \frac{[R(z)-1]H'(V)H(U)}{[H(U)+R(z)H(V)]^2} \\ &+ \frac{R'(z)H(V)[H(U)+H(V)]}{[H(U)+R(z)H(V)]^2}, \end{aligned} \tag{24}$$

where $R'(z)$ and $H'(V)$ are the derivatives with respect to z of the redundancy measure and entropy of the system output. With a small number of segments, $R(z) \approx 1$ and $H(V) \ll H(U)$. In this case, (23) implies that $VI_n(U,V) \approx 1$. For $z < z_{\min}$, both terms in expression (24) are negative. Hence, the value of the normalized variation of information decreases. Then, $|1-R(z)| < |1-R(z_{\min})|$, $\dfrac{H(U)}{H(V(z))} + R(z) > \dfrac{H(U)}{H(V(z_{\min}))} + R(z_{\min})$, and

$$\left|\frac{[1-R(z)]}{\dfrac{H(U)}{H(V(z))} + R(z)}\right| < \left|\frac{[1-R(z_{\min})]}{\dfrac{H(U)}{H(V(z_{\min}))} + R(z_{\min})}\right|,$$

which implies that $VI_n(U,V(z_{\min})) < VI_n(U,V(z))$ for $z < z_{\min}$.

At the point corresponding to $R(z_{\min})$, taking into account (9), expression (24) implies

$$\begin{aligned} VI_n'(U,V) &= \frac{[R(z_{\min})-1]H'(V(z_{\min}))H(U)}{[H(U)+H(V(z_{\min}))R(z_{\min})]^2} \\ &= \frac{[R(z_{\min})-1]H(U)}{z_{\min}[H(U)+R(z_{\min})\log z_{\min}]^2} < 0. \end{aligned}$$

For $z > z_{\min}$, $R(z) > R(z_{\min})$ and the first term in expression (24) satisfies the following inequality:

$$\left|\frac{[R(z)-1]H(U)}{z[H(U)+R(z)\log z]^2}\right| < \left|\frac{[R(z_{\min})-1]H(U)}{z_{\min}[H(U)+R(z_{\min})\log z_{\min}]^2}\right|.$$

In addition, at $z > z_{\min}$, the second term on the right-hand side of expression (24) is positive and increases with increasing z, whereas the absolute value of the first term decreases. Therefore, the rate of decrease in $VI_n(U,V)$ at $z \geq z_{\min}$ drops. This indicates that the segmented image corresponding to the minimum of information redundancy has a sufficiently small informational difference in terms of measure (22) from the image input to segmentation system (3). Graphs that characterize the dependence of measure $VI_n(U,V)$ on the number of image (V) segments at different values of parameter c in function (12) of the segmentation model are shown in Fig. 7.

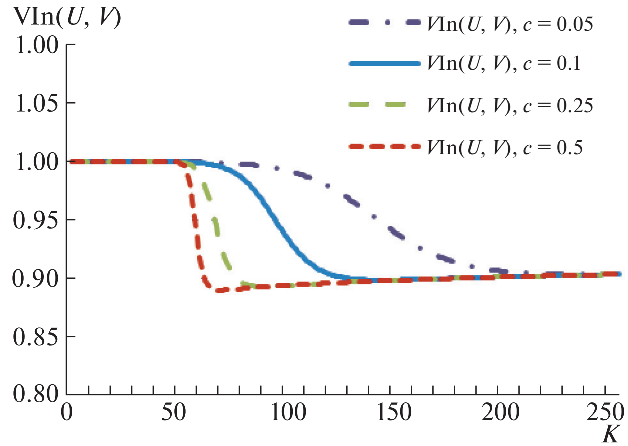

Fig. 7. Measure of difference $VI_n(U,V)$ vs. the number of segments for image V at different values of parameter c.

The effectiveness of the information redundancy criterion in image segmentation is confirmed by the computational experiment described in the next section.

2. COMPUTATIONAL EXPERIMENT

The computational experiment that is carried out to estimate the effectiveness of the criterion consists of

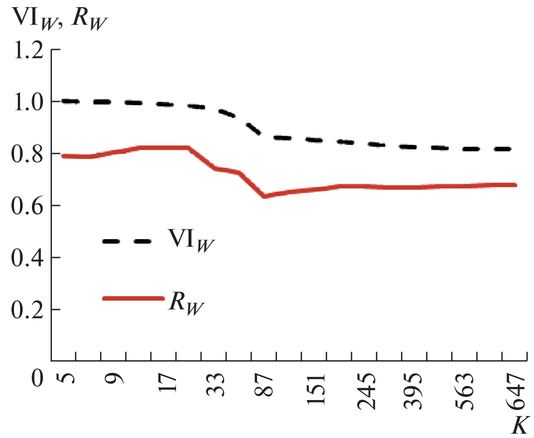

Fig. 8. Weighted redundancy measure R_w and normalized variation of information $VI_w(U,V_q)$ vs. the number of segments for the color version of the image from Fig. 1a.

two stages. At the first stage, the criterion of minimum information redundancy is used in combination with the simple linear iterative clustering (SLIC) algorithm [1] for segmenting images from BSDS500 dataset [2]. At the second stage, the results of image segmentation based on the criterion of minimum information redundancy and the entropy criterion proposed in [29] are compared.

2.1. Using the Criterion of Minimum Information Redundancy in Combination with the SLIC Algorithm

In this work, to demonstrate the validity of the proposed image segmentation model, we use images from BSDS500 [2]. Each of the analyzed images is segmented using a modified SLIC algorithm [1] at different values of postprocessing parameter Δ_1 [24]. As a result of segmenting image U, we obtain a set of Q images $\mathfrak{V} = \{V_1, V_2, \dots, V_Q\}$. For image U and each of images V_q, $q = 1,2,\dots,Q$ obtained at different values of the image segmentation parameter, redundancy measure R is evaluated and image $V_{R\min}$ corresponding to the global minimum of R is found among images V_q. To take into account all the color channels of the image, a weighted redundancy measure is used:

$$R_w(U,V_q) = \frac{R_L H_L(U) + R_a H_a(U) + R_b H_b(U)}{H_L(U) + H_a(U) + H_b(U)}, \quad (27)$$

where R_i is the redundancy measure computed for channel $i \in \{L,a,b\}$ of the CIELAB color space for images U and V_q, while H_i is the entropy of color channel i for image U. We select image $V_{q\min}$ that minimizes measure R_w: $R_w(U,V_{q\min}) = R_{\min}$. The results of this stage of the experiment are illustrated by a color version of the image (35010.jpg) shown in Fig. 1a. The values of the weighted redundancy measure for this image are represented by the solid line in Fig. 8. The minimum of the redundancy measure is reached at $K = 87$.

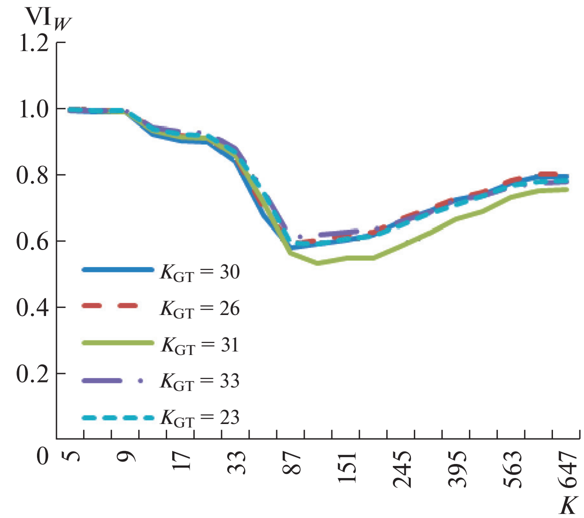

Fig. 9. Normalized variation of information $VI_w(U_i^{GT},V_q)$, $i = 1,2,\dots,5$, vs. the number of segments when comparing segmented images V_q with ground truth segmentations U_i^{GT} from BSDS500.

Then, segmented images V_q, $q = 1,2,\dots,Q$, are compared with original image U and ground truth segmentations V_s^{GT}, $t = 1,2,\dots,S$ from BSDS500 database. For this purpose, we use the weighted normalized variation of information [3, 22]:

$$VI_w(U,V_q) = \frac{VI_{nL}H_L(U) + VI_{na}H_a(U) + VI_{nb}H_b(U)}{H_L(U) + H_a(U) + H_b(U)}, \quad (28)$$

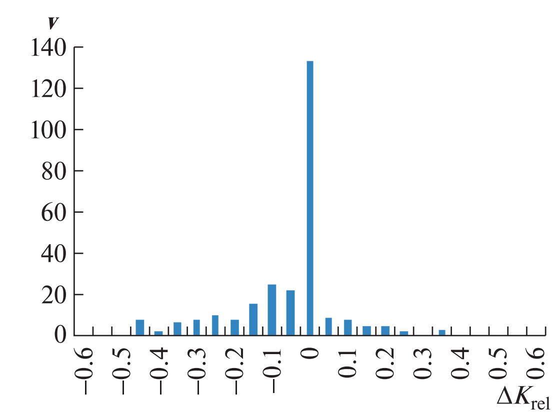

Fig. 10. Histogram for ΔK_{rel} computed when segmenting images from BSDS500, where ν is the frequency of occurrence of a certain value of ΔK_{rel}.

Fig. 11. Test image: 118035.jpg.

$$VI_{ni}(U,V_q) = \frac{H_i(U) + H_i(V_q) - 2I_i(U;V_q)}{H_i(U,V_q)}, \quad (29)$$

where $VI_w(U,V_q)$ is the weighted variation of information, VI_{ni} is the normalized variation of information between color channels i of images U and V_q, $I_i(U;V_q)$ is the mutual information between color channels of the images, and $H_i(U,V_q)$ is the joint entropy in the ith color channel.

Based on the results of comparing the original image (35010.jpg, denoted by U) with images V_q obtained at $0 \le \Delta_1 \le 3.6$, Fig. 8 plots variation of information $VI_w(U,V_q)$, shown by the dashed line, against number of segments K.

The next task is to compare five ground truth segmentations of the image under analysis (corresponding numbers of image segments K_{GT} are 30, 26, 31, 33, and 23) with a series of segmented images V_q obtained by the SLIC algorithm [1] with the post-processing procedure [24] for parameter values $0 \le \Delta_1 \le 3.6$. The result of the comparison is shown in Fig. 9. Three of the five ground truth segmentations have the minimum distance, in terms of measure (28)–(29) (which implies the maximum similarity), to the image with 87 segments obtained at $\Delta_1 = 2$, for which the minimum of redundancy measure $R_w(U,V_q)$ computed by formula (27) (see Fig. 8) is reached.

To evaluate the quality of image segmentation, we use the relative difference

$$\Delta K_{\mathrm{rel}} = \frac{K_{\min} - K_{\min}^{GT}}{K_{\max}},$$

where $K_{\min}$ is the number of image segments that provides the minimum of redundancy criterion $R_w(U,V_{q\min}) = R_{\min}$, $K_{\min}^{GT}$ is the number of segments in image V_q that provides the minimum variation of information $VI_w(V_s^{GT},V_q)$ as compared to ground truth

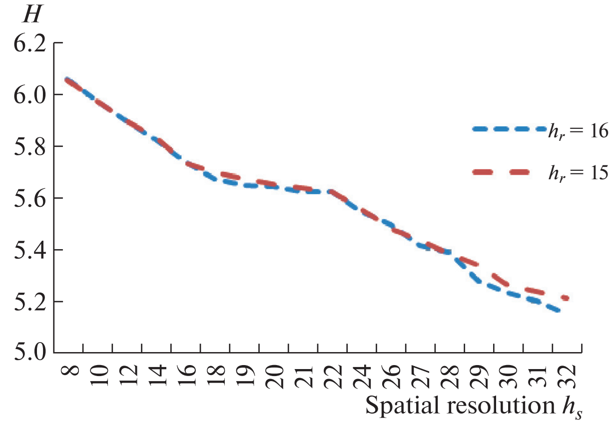

Fig. 12. Mean entropy $H_{\mathrm{mean}}(V_q)$ of the filtered image from Fig. 11 at EDISON parameters $h_r = 15$, $h_r = 16$, and $8 \le h_s \le 32$.

segmentation V_s^{GT}, and $K_{\max}$ is the maximum number of segments in images $V_1, V_2, \ldots, V_Q$. A histogram for ΔK_{rel}, which is constructed based on the results of segmenting 54 images from BSDS500 and comparison with 270 ground truth segmentations, is shown in Fig. 10. It can be seen that there is a sufficiently large group of images for which the results of segmentation by the criterion of minimum information redundancy demonstrate high informational similarity to the ground truth segmentations.

2.2. Comparison of the Proposed Method with the Method Based on the Entropy Criterion

We compared results of segmenting the images from BSDS500 by the criterion of minimum information redundancy and the entropy criterion proposed in [29]. The comparison results are presented for the image shown in Fig. 11 (118035.jpg).

For segmentation, we use the mean shift algorithm [8] implemented in the EDISON system. In accordance with the algorithm proposed in [29], using the EDISON system, we obtain a set of filtered (V_q^{filt}) and segmented (V_q, $q = 1,2,\ldots,Q$) images with different parameter values. Spatial resolution h_s varies from 8 to 32, while range resolution h_r varies from 4 to 16. The smallest significant feature size is $M = 30$. For each of the filtered images, the mean entropy is computed:

$$H_{\mathrm{mean}}(V_q^{\mathrm{filt}}) = \frac{1}{3}\sum_i H_i(V_q^{\mathrm{filt}}),$$

where H_i is the entropy of color channel i for image V_q^{filt}. Figure 12 shows the mean entropies for $h_r = 15$, $h_r = 16$, and $8 \le h_s \le 32$. When using the entropy cri-

Fig. 13. Results of segmenting the image from Fig. 11: (a) based on the criterion of minimum information redundancy, (b) based on the entropy criterion [29] with threshold *edsEnt* = 0.005, and (c) based on the entropy criterion [29] with *edsEnt* = 0.012.

terion under the condition that the difference of entropies H_{mean} of the images after filtering by the EDISON algorithm does not exceed the value $edsEnt = 0.005$ recommended in [29], we obtain the image with the number of segments $K = 36$ at $h_r = 16$ and $h_s = 22$. The value of the normalized variation of information between the input image and the image obtained by the entropy criterion ($V_{q\,\min}$) is $VI_w = 0.7897$. For $edsEnt = 0.012$, we obtain image $V_{q\,\min}$ with the number of segments $K = 30$ at $h_r = 16$ and $h_s = 20$, for which $VI_w = 0.7894$.

The minimum value of the redundancy measure for image $V_{q\,\min}$ from the obtained set of segmented images V_q, $q = 1, 2, \ldots, Q$, is $R_{w\,\min} = 0.2923$ for number of segments $K = 35$, $h_r = 15$, and $h_s = 16$. In this case, the value of the weighted normalized variation of information (28) between the original and segmented images is $VI_w(U, V_{q\,\min}) = 0.7863$. Thus, the segmented image obtained by the condition of minimum redundancy is more similar to the input image than the image obtained with the entropy criterion proposed in [29]. The images segmented by different criteria are shown in Fig. 13. The segmentation results are visually similar. In the image corresponding to the information redundancy minimum, the bell is better distinguished (see Fig. 13a), while in the images obtained by the entropy criterion, some of the small bell tower details are slightly better distinguished (see Figs. 13b, 13c).

For quantitative comparison of the results, we estimate information difference $VI_w(V_s^{GT}, V_{q\,\min})$, $s = 1, 2, \ldots, 5$, (28)–(29) for five ground truth segmentations V_s^{GT} from BSDS500 and segmented images $V_{q\,\min}$ obtained using the criteria discussed above. The comparison results are shown in Table 1. It can be seen that, in three out of five cases, the segmented image obtained by the criterion of minimum information redundancy has smaller difference from the ground truth segmentations than the segmented image obtained by the minimum entropy criterion when threshold $edsEnt = 0.012$ and in four out of five cases when $edsEnt = 0.0005$. It should be noted that, in general, the segmentation results that minimize these criteria have quite similar variation of information as compared to the ground truth segmentations from BSDS500.

CONCLUSIONS

In this paper, we have described an iterative information-theoretical method for evaluating the quality of digital image segmentation. A system has been considered that includes a segmentation algorithm with a parameter determining the number of image segments and the iterative procedure for setting the value of this parameter that provides the minimum of the quality functional. The image segmentation algorithm has been regarded as an information channel. The simplified mathematical model of the digital image segmentation algorithm has been proposed. Based on Barlow's hypothesis [5] and the principle of minimum

Table 1. Weighted normalized variation of information $VI_w(V_s^{GT}, V_{q\,\min})$, $s = 1, 2, \ldots, 5$ between the ground truth segmentations of the image from BSDS500 and the segmentations obtained using the EDISON algorithm with different criteria

Criterion	$VI_w(V_s^{GT}, V_{q\,\min})$, $s = 1, \ldots, 5$				
	Ground truth segmentation 1	Ground truth segmentation 2	Ground truth segmentation 3	Ground truth segmentation 4	Ground truth segmentation 5
min R_w	0.4060	0.4492	0.5363	0.3771	0.4353
$edsEnt \le 0.005$	0.4125	0.4551	0.5372	0.3423	0.4419
$edsEnt \le 0.012$	0.4065	0.4471	0.5376	0.3311	0.4390

redundancy in the visual perception model [4], the segmentation quality index—the information redundancy measure—has been selected. It has been shown that, for the proposed model, there exists a minimum of the redundancy measure, which corresponds to the best image partition. The relationships among the model parameters for which this minimum is reached have been derived. The validity of the model has been confirmed by the computational experiment on images from the Berkeley Segmentation Dataset (BSDS500). It has been found that, for a sufficiently large group of test images, the segmentation results obtained by the condition of minimum redundancy have the minimum variation of information as compared to the ground truth segmentations from BSDS500. With the ground truth images being manually segmented by specialists, it can be concluded that the segmentations obtained based on the condition of minimum information redundancy are the best segmentations in terms of visual perception. This does not contradict Barlow's hypothesis [5] that information redundancy is minimized at the initial stages of signal processing in the human visual system.

We compared the results of image segmentation by the EDISON system based on the criterion of minimum information redundancy and the entropy criterion. The segmented image obtained based on the condition of minimum redundancy is more similar to the input image than the image obtained by the entropy criterion, and, in most cases, it is more similar to the ground truth segmentations.

In future works, we intend to carry out a deeper investigation of the developed model and properties of the segmentation quality index based on the redundancy measure.

FUNDING

This research was carried out at the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences in the framework of a state task.

COMPLIANCE WITH ETHICAL STANDARDS

This paper is a completely original work of its author, it has not been published before, and it will not be sent to other publishers until the decision of the *PRIA* Editorial Board is received.

Conflict of Interest

The process of writing and the content of the article does not give grounds for raising the issue of a conflict of interest.

REFERENCES

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2274–2282 (2012). https://doi.org/10.1109/tpami.2012.120
2. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. on Computer Vision (ICCV 2001), Vancouver, 2002*, (IEEE, 2001), Vol. 2, pp. 416–423. https://doi.org/10.1109/ICCV.2001.937655
3. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *IEEE Conf. on Computer Vision and Pattern Recognition, Miami, 2009* (IEEE, 2009), pp. 2294–2301. https://doi.org/10.1109/CVPR.2009.5206707
4. J. J. Atick and A. N. Redlich, "Towards a theory of early visual processing", Neural Comput. **2**, 308–320 (1990). https://doi.org/10.1162/neco.1990.2.3.308
5. H. B. Barlow, "Possible principles underlying the transformations of sensory messages". Sens. Commun., **1**, 217–234 (1961). https://doi.org/10.7551/mitpress/9780262518420.003.0013
6. R. P. Bohush, S. V. Ablameyko, E. R. Adamovskiy, and D. Savca, "Image similarity estimation based on ratio and distance calculation between features," Pattern Recognit. Image Anal. **30**, 147–159 (2020). https://doi.org/10.1134/S1054661820020030
7. M. Borsotti, P. Campadelli, and R. Schettini, "Quantitative evaluation of color image segmentation results", Pattern Recognit. Lett. **19**, 741–747 (1998). https://doi.org/10.1016/S0167-8655(98)00052-X
8. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603–619 (2002). https://doi.org/10.1109/34.1000236
9. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation", Int. J. Comput. Vision, **59**, 167–181 (2004). https://doi.org/10.1023/B:VISI.0000022288.19776.77
10. E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings", J. Am. Stat. Assoc. **78**, 553–569 (1983).
11. A. L. N. Fred and A. K. Jain, "Robust data clustering," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Madison, Wis., 2013* (IEEE, 2003), Vol. 2, pp. 128–133. https://doi.org/10.1109/CVPR.2003.1211462
12. A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," IEEE Trans. Pattern Anal. Mach. Intell. **27**, 835–850 (2005). https://doi.org/10.1109/TPAMI.2005.113
13. I. Frosio and E. R. Ratner, "Adaptive segmentation based on a learned quality metric", in *Proc. 10th Int. Conf. on Computer Vision Theory and Applications*, Vol. 2: *VISAPP* (SciTePress, 2015), pp. 283–292. https://doi.org/10.5220/0005257202830292
14. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Pearson Prentice Hall, Upper Saddle River, N.J., 2008).
15. R. M. Haralick and L. G. Shapiro, "Image segmentation techniques", Comput. Vision, Graphics, and Image Process. **29**, 100–132 (1985). https://doi.org/10.1016/S0734-189X(85)90153-7

16. L. Khelifi and M. Mignotte, "EFA-BMFM: A Multi-criterion framework for the fusion of colour image segmentation," Inf. Fusion **38**, 104–121 (2017). https://doi.org/10.1016/j.inffus.2017.03.001
17. T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, and L. Grady, "Evaluating segmentation error without ground truth," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, Ed. by N. Ayache, H. Delingette, P. Golland, and K. Mori, Lecture Notes in Computer Science, vol. 7510 (Springer, Berlin, 2012), pp. 528–536. https://doi.org/10.1007/978-3-642-33415-3_65
18. J. Liu and Y.-H. Yang, "Multiresolution color image segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 689–700 (1994). https://doi.org/10.1109/34.297949
19. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", in *Proc. Eighth IEEE Int. Conf. on Computer Vision (ICCV 2001), Vancouver, 2001* (IEEE, 2001), Vol. 2, pp. 416–423. https://doi.org/10.1109/ICCV.2001.937655
20. D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues", IEEE Trans. Pattern Anal. Mach. Intell., **26**, 530–549 (2004). https://doi.org/10.1109/TPAMI.2004.1273918
21. M. Meilă, "Comparing clusterings: an axiomatic view," in *Proc. 22nd Int. Conf. on Machine Learning, Bonn, Germany, 2005* (Assoc. for Computing Machinery, New York, 2005), pp. 577–584. https://doi.org/10.1145/1102351.1102424
22. M. Mignotte, "A label field fusion model with a variation of information estimator for image segmentation. information fusion", Inf. Fusion **20**, 7–20 (2014). https://doi.org/10.1016/j.inffus.2013.10.012
23. D. M. Murashov, "Localization of differences between multimodal images on the basis of an information-theoretical measure," Pattern Recognit. Image Anal. **24**, 133–143 (2014). https://doi.org/10.1134/S105466181401012X
24. D. Murashov, "Theoretical-information quality model for image segmentation", Procedia Eng. **201**, 239–248 (2017). https://doi.org/10.1016/j.proeng.2017.09.603
25. D. Murashov, "Application of information redundancy measure to image segmentation," in *Intelligent Data Processing. IDP 2016*, Ed. by V. Strijov, D. Ignatov, and K. Vorontsov, Communications in Computer and Information Science, vol. 794 (Springer, Cham, 2019), pp. 125–139. https://doi.org/10.1007/978-3-030-35400-8_9
26. W. M. Rand, "Objective criteria for the evaluation of clustering methods", J. Am. Stat. Assoc. **66**, 846–850 (1971). https://doi.org/10.1080/01621459.1971.10482356
27. J. Rigau, M. Feixas, M. Sbert, A. Bardera, and I. Boada, "Medical Image Segmentation Based on Mutual Information Maximization", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, Ed. by C. Barillot, D. R. Haynor, and P. Hellier, Lecture Notes in Computer Science, vol. 3216 (Springer, Berlin, 2004), pp. 135–142. https://doi.org/10.1007/978-3-540-30135-6_17
28. R. Robinson, O. Oktay, W. Bai, V. V. Valindria, M. M. Sanghvi, N. Aung, J. M. Paiva, F. Zemrak, K. Fung, E. Lukaschuk, A. M. Lee, V. Carapella, Y. Jin Kim, B. Kainz, S. K. Piechnik, S. Neubauer, S. E. Petersen, C. Page, D. Rueckert, and B. Glocker, "Real-time prediction of segmentation quality", in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Ed. by A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Lecture Notes in Computer Science, vol. 11073 (Springer, Cham, 2018), pp. 578–585. https://doi.org/10.1007/978-3-030-00937-3_66
29. R. Rodríguez and A. G. Suarez, "An image segmentation algorithm using iteratively the mean shift," in *Progress in Pattern Recognition, Image Analysis and Applications. CIARP 2006*, Ed. by J. F. Martínez-Trinidad, J. A. Carrasco Ochoa, and J. Kittler, Lecture Notes in Computer Science, vol. 4225 (Springer, Berlin, 2006), pp. 326–335. https://doi.org/10.1007/11892755_33
30. A. Strehl and J. Ghosh, "Cluster ensembles – A knowledge reuse framework for combining multiple partitions," J. Mach. Learning Res. **3**, 583–617 (2002). https://doi.org/10.1162/153244303321897735
31. R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, San Diego, Calif., 2005* (IEEE, 2005), pp. 34–41. https://doi.org/10.1109/CVPR.2005.390
32. H. Zhang, J. E. Fritts, and S. A. Goldman, "An entropy-based objective evaluation method for image segmentation," Proc. SPIE **5307**, 38–49 (2003). https://doi.org/10.1117/12.527167
33. H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," Comput. Vision Image Understanding **110**, 260–280 (2008). https://doi.org/10.1016/j.cviu.2007.08.003

Translated by Yu. Kornienko

Dmitry Mikhailovich Murashov. Born January 30, 1958. Graduated from the Ordzhonikidze Moscow Aviation Institute in 1981 (with the specialization of Automatic Control Systems). Received his Candidate's degree in 1990. Associate Professor in the specialty of Theoretical Foundations of Informatics. He currently works at the Institute of Cybernetics and Educational Informatics of the Federal Research Center "Computer Science and Control" (Russian Academy of Sciences), Moscow. Author of more than 80 papers. Scientific interests: automatic control, image processing and analysis, and pattern recognition.