Classification of Sign-Based Image Representations Based on Distance Functions

A. G. Bronevich[a] and A. V. Goncharov[b]

[a]*National Research University Higher School of Economics, Pokrovskii bul. 11, Moscow, 109028 Russia*

[b]*CVisionLab, LLC, Severnaya Ploshchad' 3, Taganrog, 347900 Russia*

e-mail: brone@mail.ru, goncharov@cvisionlab.com

Abstract—The paper proposes various approaches to classifying sign-based representations of images based on distance functions. Any image is represented as a set of features describing differences in brightness. The construction of a distance function is proposed using classical functionals of information theory: the Shannon entropy and the Kullback–Leibler distance. It is shown that the Bayes classification in the case of independent features can be also described by distance functions. In the last section, the proposed approaches are evaluated using a face detection problem.

Keywords: sign-based representation of images, measure, informativeness, Kullback–Leibler distance.

DOI: 10.1134/S1054661813020053

INTRODUCTION

One of the basic challenges in image recognition is the dependence of the recognition results on the image recording conditions. To increase recognition performance with respect to image acquisition conditions, sign-based image representation can be used [5], which is uniquely specified by the quasi-order relation; the latter describes differences in brightness of neighboring pixels. Sign-based representation is invariant under monotonically increasing brightness transformations and may be useful in solving face recognition problems [2, 3] and retrieving of near-duplicate images in large collections [4]. A theoretical study of properties of sign-based representations is given in [5–7]. It is assumed that the problem of classifying sign-based representations of images can be solved by introducing distance functions. When studying analysis of expert information from a theoretical viewpoint, B.G. Litvak proposed an axiomatic approach to constructing metrics with the aid of partial order relation; this approach leads to the Hamming metric or the Euclidean metric [10].

In this paper, we propose a new approach to introducing a distance function on sign-based representations. To do so we introduce descriptions of sign-based representations by using a finite set of symbolic features; from here on, the distance function is defined as the amount of information lost in assuming that the compared sign-based representations coincide. In a particular case, this approach leads to the Hamming metric, which was constructed by Litvak in [10].

Received October 19, 2010

1. SIGN-BASED REPRESENTATION OF IMAGES

A nonnegative integer function $f(\mathbf{x})$, $\mathbf{x} = (x_1, x_2)$ given at nodes of a grid $\Omega = I_N \times I_M = \{1, \ldots, N\} \times \{1, \ldots, M\}$ is called an image. The set of all images $f: \Omega \longrightarrow \mathbb{Z}_+$ is denoted by $\mathcal{F}$.

Definition 1. *A relation $\tau \subseteq \Omega \times \Omega$ is called a sign-based representation of an image $f \in \mathcal{F}$ provided that the following conditions are satisfied* [3]:

(1) if $(\mathbf{x}_1, \mathbf{x}_2) \in \tau$, then $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$;

(2) if $(\mathbf{x}_1, \mathbf{x}_2) \in \tau$, $(\mathbf{x}_2, \mathbf{x}_1) \notin \tau$, then $f(\mathbf{x}_1) < f(\mathbf{x}_2)$.

As examples we consider complete and neighborhood sign-based representations [6, 7]. A sign-based representation is called *complete* if at least one of the pairs $(\mathbf{x}_1, \mathbf{x}_2)$ or $(\mathbf{x}_2, \mathbf{x}_1)$ $\mathbf{x}_2 \ne \mathbf{x}_1$ lies in τ:

$$\tau = \{(\mathbf{x}_1, \mathbf{x}_2) \in \Omega^2 \,|\, f(\mathbf{x}_1) \le f(\mathbf{x}_2), \mathbf{x}_1 \ne \mathbf{x}_2\}.$$

If relation τ contains only pairs of adjacent pixels in the sense of neighborhood $O_\varepsilon(\mathbf{x}_1) = \{\mathbf{x}_2 \in \Omega \,|\, \|\mathbf{x}_1 - \mathbf{x}_2\| \le \varepsilon\}$, where $\|\mathbf{x}_1 - \mathbf{x}_2\| = |x_1 - y_1| + |x_2 - y_2|$, then τ is a *neighborhood sign-based representation*:

$$\tau = \{(\mathbf{x}_1, \mathbf{x}_2) \in \Omega^2 \,|\, f(\mathbf{x}_1) \le f(\mathbf{x}_2), \mathbf{x}_2 \in O_\varepsilon(\mathbf{x}_1)\}.$$

Let us notice that any sign-based representation τ defines on the set $\mathcal{F}$ a set of images with this sign-based representation. We denote this class of images as C_τ. From Definition 1 it follows that C_τ can be regarded as a solution of the following system of linear inequalities:

$$\begin{cases} f(\mathbf{x}_i) < f(\mathbf{x}_j) & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in \tau, \ (\mathbf{x}_j, \mathbf{x}_i) \notin \tau, \\ f(\mathbf{x}_i) = f(\mathbf{x}_j) & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in \tau, \ (\mathbf{x}_j, \mathbf{x}_i) \in \tau. \end{cases}$$

This system contains an inequality (equalities) for all pairs of pixels $(\mathbf{x}_i, \mathbf{x}_j) \in \tau$. One may also consider the search problem of all sign-based representations that generate the same class of images. It is easily shown that $C_{\tau_1} = C_{\tau_2}$ for sign-based representations τ_1 and τ_2 if and only if $(\tau_1)^{\mathrm{Tr}} = (\tau_2)^{\mathrm{Tr}}$, where "Tr" denotes the transitive closure of a relation. So, two sign-based representations τ_1 and τ_2 is considered to be *equivalent* if $C_{\tau_1} = C_{\tau_2}$ or $(\tau_1)^{\mathrm{Tr}} = (\tau_2)^{\mathrm{Tr}}$. In the latter case, any sign-based representation τ is uniquely described by the quasi-order relation τ^{Tr}. Some subsequent results in this direction may be found in [6], which contains, in particular, necessary and sufficient conditions that a relation τ be a neighborhood sign-based representation of a image.

2. DISTANCE FUNCTIONS ON SIGN-BASED REPRESENTATIONS

The application of a special kind of distance functions was validated by many researches in the field of computer vision. In particular, a special closeness measure on images is proposed in [8]; it is based on coefficients of wavelet transformations for which the symmetry property and the triangle inequality do not hold. Paper [9] concerns a face detection method based on application of the Hausdorff distance for comparison of images with a face pattern; this function does not have the symmetry property.

Let us introduce the concepts of distance function and metric on sign-based representations.

Definition 2. *A distance function on sign-based representations is a mapping* d: $\mathcal{F} \times \mathcal{F} \longrightarrow [0, \infty)$ *satisfying the following properties*:

1. $d(f_1, f_2) = d(f_2, f_1)$ *for all* $f_1, f_2 \in \mathcal{F}$;

2. $d(f_1, f_2) = d(f_1, f_3)$ if $d(f_2, f_3) = 0$ *for all* $f_1, f_2, f_3 \in \mathcal{F}$;

3. to each equivalence class $C_f = \{g \in \mathcal{F} | d(f, g) = 0\}$ *there corresponds some sign-based representation of an image; i.e., there exists a sign-based representation of* τ *such that* $C_\tau = C_f$.

A particular case of a distance function is a metric on sign-based representations, which will be defined later.

Definition 3. *A metric on sign-based representations is a distance function* d: $\mathcal{F} \times \mathcal{F} \longrightarrow [0, \infty)$ *such that* $d(f_1, f_2) + d(f_2, f_3) \ge d(f_1, f_3)$ *for all* $f_1, f_2, f_3 \in \mathcal{F}$.

The triangle inequality $d(f_1, f_2) + d(f_2, f_3) \ge d(f_1, f_3)$, which is necessary for a metric, implies property (2) of a distance function. Indeed, let us assume that $d(f_2, f_3) = 0$; then from the inequalities $d(f_1, f_2) + d(f_2, f_3) \ge d(f_1, f_3)$, $d(f_1, f_3) + d(f_3, f_2) \ge d(f_1, f_2)$ imply $d(f_1, f_2) = d(f_1, f_3)$.

Assume that $f_1, f_2 \in C_f$ and $g \in \mathcal{F}$. Then from property (2) we see that $d(f_1, g) = d(f_2, g)$. So, we may assume that a metric is introduced on the classes of equivalence and $d(C_f, C_g) = d(f, g)$ for $f, g \in \mathcal{F}$

Definition 2 also implies that the distance function on sign-based representations determines some partition of set $\mathcal{F}$ into disjoint classes to which sign-based representations correspond. In other words, a metrics cannot be introduced on all possible sign-based representations, but only on some subset of sign-based representations that determine a partition of set $\mathcal{F}$. (In other words, a metrics may be introduced on some subset of sign-based representations only, that determine a partition of set $\mathcal{F}$.) Let us see how these partitions are generated.

Let τ be a sign-based representation of an image f and $C_\tau \subset \mathcal{F}$ be a class of images corresponding to a sign-based representation τ. Then we shall consider elementary sign-based representations to which correspond equivalence classes of the form $C_{i,j,0} = \{f \in \mathcal{F} | f(\mathbf{x}_i) = f(\mathbf{x}_j)\}$ and $C_{i,j,1} = \{f \in \mathcal{F} | f(\mathbf{x}_i) < f(\mathbf{x}_j)\}$ for $i \ne j$. Clearly, these classes correspond to elementary quasi-order relations that provide information only on intensity values of pixels $\mathbf{x}_i$ and $\mathbf{x}_j$. The classes $C_{i,j,0}$, $C_{i,j,1}$, and $C_{j,i,1}$ for $i \ne j$ are elementary partitions of set $\mathcal{F}$; each class C_τ can be represented as a finite intersection of elementary classes. In order to obtain such a representation, it is necessary to describe the class C_τ by a system of linear inequalities: $f(\mathbf{x}_i) < f(\mathbf{x}_j)$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \tau$, $(\mathbf{x}_j, \mathbf{x}_i) \notin \tau$; $f(\mathbf{x}_i) = f(\mathbf{x}_j)$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \tau$, $(\mathbf{x}_j, \mathbf{x}_i) \in \tau$

$$C_\tau = \bigcap_{\substack{(\mathbf{x}_i, \mathbf{x}_j) \in \tau \\ (\mathbf{x}_j, \mathbf{x}_i) \in \tau}} C_{i,j,0} \cap \bigcap_{\substack{(\mathbf{x}_i, \mathbf{x}_j) \in \tau \\ (\mathbf{x}_j, \mathbf{x}_i) \notin \tau}} C_{i,j,1}.$$

Thus a sign-based representation τ of image f can be regarded as a set of images $C_\tau \subset \mathcal{F}$ defined by the relation τ.

We point out that the sign of the expression $f(\mathbf{x}_j) - f(\mathbf{x}_i)$ can be regarded as some feature:

$$\pi(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} -1, & f(\mathbf{x}_j) < f(\mathbf{x}_i), \\ 0, & f(\mathbf{x}_j) = f(\mathbf{x}_i), \\ 1, & f(\mathbf{x}_j) > f(\mathbf{x}_i). \end{cases} \tag{2.1}$$

As a rule, a decision about membership of an image to some class is made from the set of features. Taking this into consideration, we shall assume below that we have some set of features described by an antisymmetric and antireflexive relation with $\alpha \in \Omega \times \Omega$, the features being calculated for each pair of pixels $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$. In this case, as a partition of the set $\mathcal{F}$ we can take all nonempty subsets of the form

$$C(\alpha_1, \alpha_2, \alpha_3) = \bigcap_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha_1} C_{i,j,0}$$

$$\cap \bigcap_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha_2} C_{i,j,1} \cap \bigcap_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha_3} C_{j,i,1},$$

where $\alpha_1 \cup \alpha_2 \cup \alpha_3 = \alpha$ and α_i, $i = 1, 2, 3$, are pairwise disjoint sets. To choose the distance function on sign-based representations, we consider an approach based on measuring the amount of information. We assume that, for any feature $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$, its informativeness $w(\mathbf{x}_i, \mathbf{x}_j) > 0$ can be specified (it is assumed that set α is chosen so that the informativeness for each feature is positive). Then, as the value of the distance between sign-based representations of images $f, g \in \mathcal{F}$, we can take the amount of information lost in assuming that the sign-based representations τ_f and τ_g are equal.

Let $\alpha^*(f, g) = \{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha | \pi_f(\mathbf{x}_i, \mathbf{x}_j) \neq \pi_g(\mathbf{x}_i, \mathbf{x}_j)\}$ be a subset of pairs of pixels on which the sign-based representations τ_f and τ_g are different.

In the first model, we assume that the features are independent. Then,

$$d_H(f, g) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha^*(f, g)} w(\mathbf{x}_i, \mathbf{x}_j).$$

The independence of features is a rather strict assumption, which does not necessarily hold in practice. For complete sign-based representations, it is clear that the independence condition is violated, because features are dependent due to the transitivity conditions of the relations in question.

We assume that to any subset of features $A \subseteq \alpha$ there corresponds an informativeness value $\mu(A) \geq 0$ that satisfies the following properties of a nonadditive measure [13]:

(a) $\mu(\varnothing) = 0$;

(b) $\mu(A) \leq \mu(B)$ if $A \subseteq B$.

Then the distance function between sign-based representations can be defined as follows:

$$d_\mu(f, g) = \mu(\alpha^*(f, g)). \quad (2.2)$$

The following proposition gives necessary and sufficient conditions that a function d_μ be a metric on sign-based representations.

Proposition 1. *The function d_μ is a metric on sign-based representations defined by a system of features $\alpha \subseteq \Omega \times \Omega$ if and only if*

(1) $\mu(A) > 0$ *for all* $A \subseteq \alpha$ *for* $|A| > 0$;

(2) $\mu(A) + \mu(B) \geq \mu(A \cup B)$ *for all* $A, B \subseteq \alpha$ *for* $A \cap B = \varnothing$.

Proof. Clearly, condition (1) is necessary, since the value of function d_μ must be greater than zero for sign-based representations that differ from each other by a value of at least one feature. Inequality (2) follows from the triangle inequality. Indeed, consider sign-based representations τ_1, τ_2, τ_3 of images $f_1, f_2, f_3 \in \mathcal{F}$. Assume that τ_1 and τ_3 differ from each other by a set of features $C \subseteq \alpha$. If τ_1 and τ_2 differ from each other by a set of features $A \subseteq \alpha$, and if τ_2 and τ_3 differ from each other by a set of features $B \subseteq \alpha$, then, clearly, $C \subseteq A \cup B$. So, the following inequality must hold in view of the triangle inequality: $\mu(A) + \mu(B) \geq \mu(C)$ for any $A, B, C \subseteq \alpha$ with $C \subseteq A \cup B$. Due to the monotonicity of the nonadditive measure μ, this inequality is equivalent to the following: $\mu(A) + \mu(B) \geq \mu(A \cup B)$ for all $A, B \subseteq \alpha$ for $A \cap B = \varnothing$.

Remark 1. *It is noteworthy that condition (1) of Proposition 1 in fact means that we should choose an informative system of features. Condition (2) is the so-called subadditivity property of information measure; its fulfillment was validated by many researches in information theory. In particular, if $\mu(A) + \mu(B) = \mu(A \cup B)$ for all $A, B \in \alpha$ with $A \cap B = 0$, then the measure μ is additive; this corresponds to the case of independence of features.*

Remark 2. *Assume that condition (1) of Proposition 1 is not satisfied. Then there exists a feature $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$ such that $\mu(\{(\mathbf{x}_i, \mathbf{x}_j)\}) = 0$. Note that the subadditivity condition μ implies that $\mu(A \backslash \{(\mathbf{x}_i, \mathbf{x}_j)\}) = \mu(A)$ for any $A \subseteq \alpha$. Taking this into account, noninformative features can be eliminated from consideration. It also implies that if condition (1) is not satisfied, then d_μ is also a metric on sign-based representations generated by a narrower system of features.*

In the simplest case, we can assume that features are independent and the informativeness of all features is the same; i.e., the information measure $\mu(A) = |A|$ equals the cardinality of set A (this is clearly an additive measure). In this case, metrics (2.2) is the Hamming metrics:

$$d(f, g) = |\alpha^*(f, g)|. \quad (2.3)$$

Note that metrics (2.3) on sign-based representations, as constructed using our approach, agrees with Litvak's metrics based on the axiomatic approach [10]; the latter metrics was introduced on partial order relations for examining problems related to expert information processing.

Let us now consider constructing, first, a metrics based on the Shannon entropy, and second, a distance function based on the relative entropy.

2.1. Metric Based on the Shannon Entropy

Analysis of image collections involves the possibility that each image can be simulated as a realization of a multivariate random variable f. It also assumes that features of images are random variables and so the Shannon entropy S may be useful in estimating their informativity. In this case, the information measure of a set of features $A \subseteq \alpha$ can be defined as the entropy of a random vector $(\pi_f(\mathbf{x}_i, \mathbf{x}_j))_{(\mathbf{x}_i, \mathbf{x}_j) \in A}$; namely, $\mu_S(A) = S((\pi_f(\mathbf{x}_i, \mathbf{x}_j))_{(\mathbf{x}_i, \mathbf{x}_j) \in A})$.

It is easily verified that the set function μ_S satisfies all the properties from Proposition 1. In addition, if features $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$, $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$ are independent, then

$$\mu_S(A) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in A} S(\pi_f(\mathbf{x}_i, \mathbf{x}_j)). \quad (2.4)$$

2.2. Distance Function Based on Relative Entropy

The above conventional entropy-based metric (2.4) does not take into account the differences between the classes Cl_l, $l = 1, \ldots, L$. In constructing the classification rule, it would be more advantageous to assign larger informativeness to features in which class Cl_l differs from the others.

Assume that Ω is a finite set, functions $p, q: \Omega \longrightarrow [0, 1]$ describe probability distributions on Ω, and distribution p is absolutely continuous with respect to q; i.e., the condition $q(x) = 0$ implies that $p(x) = 0$ for all $x \in \Omega$. Then the functional $S_{KL}(p, q) = \sum_{x \in \Omega} \ln((p(x)/q(x)))p(x)$ is called the entropy[1] of a distribution p with respect to the distribution q.

The relative entropy may be useful as the information measure of features. We shall denote by $\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j)$, $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$, the random variable that describes the observation of a feature $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$ in case we analyze an image from the class Cl_l, $l = 1, \ldots, k$. Consequently, the informativeness of features $A \subseteq \alpha$ that is capable of taking into account the differences in distribution of features over the classes F and G can be estimated using the following functionals:

$$\mu_{F,G}(A) = S_{KL}(\{\pi_{f|F}(\mathbf{x}_i, \mathbf{x}_j)\}_{(\mathbf{x}_i, \mathbf{x}_j) \in A}, \{\pi_{f|G}(\mathbf{x}_i, \mathbf{x}_j)\}_{(\mathbf{x}_i, \mathbf{x}_j) \in A});$$

where F and G are different classes from Cl_l, $\overline{\mathrm{Cl}_l} = \bigcup_{m|m \neq l} \mathrm{Cl}_m$ and $\mathrm{Cl} = \bigcup_{l=1}^{L} \mathrm{Cl}_l$.

Let us discuss the particularities of the introduced set functions and the corresponding distance functions. The distance functions constructed from the nonadditive measures $\mu_{\mathrm{Cl}_l, \overline{\mathrm{Cl}_l}}$ and $\mu_{\overline{\mathrm{Cl}_l}, \mathrm{Cl}_l}$ are the most sensitive to differences between classes. A distance function that is constructed from measure $\mu_{\mathrm{Cl}_l, \mathrm{Cl}}$ must be more regular. In particular, the probability distribution of features that corresponds to class Cl_l is always absolutely continuous with respect to the probability distribution of features that corresponds to the Cl; hence $\mu_{\mathrm{Cl}_l, \mathrm{Cl}}$ is always finite.

In general, the distance functions introduced above are not metrics, inasmuch as the above nonadditive measures are superadditive rather than subadditive. Indeed, consider a nonadditive measure $\mu_{F,G}$, where F and G are some classes; we also assume that features corresponding to the class G are independent. In this case, the relative entropy is superadditive [12], and $\mu_{F,G}(A) + \mu_{F,G}(B) \leq \mu_{F,G}(A \cup B)$ for $A \cap B = \varnothing$ and $A, B \subseteq \alpha$. If features from both classes F and G are independent, then $\mu_{F,G}$ is an additive measure on subsets of α, and hence, in this case the expression for $d_{F,G}$ is as follows:

$$d_{F,G}(f, g) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha^*(f,g)} \mu_{F,G}(\{(\mathbf{x}_i, \mathbf{x}_j)\}). \quad (2.5)$$

In the case in question $d_{F,G}$ is a metric on sign-based representations.

Superadditivity of a nonadditive measure follows as a corollary from the above approach to constructing distance functions. This approach is as follows. We introduce a nonnegative function $v(p^{(1)}, p^{(2)})$ on probability distributions that are considered in one measurable space, in particular, on all subsets of the set Ω. We have $v(p^{(1)}, p^{(2)}) = 0$ if and only if $p^{(1)} = p^{(2)}$. Assume that $\Omega = X \times Y$ and $p_X^{(i)}$, $p_Y^{(i)}$ are marginal distributions of $p^{(i)}$, $i = 1, 2$. Using similar arguments as in the construction of nonadditive information measures based on the relative entropy, it can be inferred that the nonadditive measure, as constructed from v, is subadditive if the inequality $v(p_X^{(1)}, p_X^{(2)}) + v(p_Y^{(1)}, p_Y^{(2)}) \geq v(p^{(1)}, p^{(2)})$ always holds. It is known that marginal distributions do not define a joint distribution in a unique fashion; hence, there exist probability distributions $p^{(1)}$ and $p^{(2)}$ such that $p^{(1)} \neq p^{(2)}$. However, $p_X^{(1)} = p_X^{(2)}$, $p_Y^{(1)} = p_Y^{(2)}$. Hence, for the case in question, $v(p_X^{(1)}, p_X^{(2)}) = v(p_Y^{(1)}, p_Y^{(2)}) = 0$ and $v(p^{(1)}, p^{(2)}) > 0$; i.e., the subadditivity property is not satisfied, The independence of features inside a class is regarded as a rather strict requirement. A weaker assumption is the independence of features inside the entire collection of images; i.e., inside the class Cl. In this case, the metric

$$d(f, g) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha^*(f,g)} \mu_{\mathrm{Cl}_l, \mathrm{Cl}}(\{(\mathbf{x}_i, \mathbf{x}_j)\})$$

is the best lower-bound approximation of $d_{\mathrm{Cl}_l, \mathrm{Cl}}(f, g)$.

3. CLASSIFICATION OF SIGN-BASED REPRESENTATIONS ON THE BASIS OF LIKELIHOOD FUNCTIONS

In this section, we assume that it is required to divide images into $\mathrm{Cl}_1, \ldots, \mathrm{Cl}_k$ classes using a training sample; the images are taken from the images $X = \{(f_i, n_i)\}_{i=1}^{N}$ consisting of pairs of the form (f_i, n_i), where $n_i \in 1, \ldots, k$ is the number of the class in which the image f lies. Note that if the indicated sample is

[1] The relative entropy is also referred to as the Kullback–Leibler distance or divergence [11, 12].

representative, then from it, it is possible to evaluate the probability p_l of occurrence of the class Cl_l: $\hat{p}_l = |\{(f_i, n_i) \in X | n_i = l\}|/N$, and, in a similar way, the probability distribution of features $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$ inside the class Cl_l. However, it does not seem possible to obtain reliable estimates of these characteristics from real statistical data.

Assume, for example, that these characteristics are calculated using some collection of images; it is required to divide images into two classes: "faces" (first class) and "nonfaces" (second class). In this case, the occurrence frequencies of classes are not indicative of their frequencies in actual practice and so to estimate p_l, $l = 1, 2$, it is necessary to learn the recognition system under real conditions, which may vary in time depending on the aims of recognition and so on. Even more problematic is estimating the distribution of features for each class, since in this case the number of features due to the use of sign-based representations is linear and in some cases is quadratic in the number of image pixels, so it does not seems tractable to obtain acceptable training samples. Hence, it appears feasible to introduce plausible assumptions in regard to the distribution of features inside, i.e., on the dependence of features (in particular, one may assume the independence of features for windowed neighborhood sign-based representations).

Let us find out how a Bayes classifier is constructed under the assumption that the features $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$ are independent. To do so we first estimate the probability distribution of each feature from a training sample inside the class Cl_l, $l = 1, \ldots, k$. Let $\Pr\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c\}$ be the probability that the feature $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$ takes the value $c \in \{-1, 0, 1\}$ for images from the class Cl_l. Then, using the training sample X, this probability is estimated as follows:

$$\hat{\Pr}(\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c) = \frac{|\{(f, n) \in X | n = l, \pi_f(\mathbf{x}_i, \mathbf{x}_j) = c\}|}{|\{(f, n) \in X | n = l\}|}.$$

The independence of features implies that the likelihood function that an image $f \in \mathcal{F}$ lies in class Cl_l is calculated by the formula:

$$\hat{\Pr}(f \in \mathrm{Cl}_l) = \hat{p}_l \prod_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j)\}.$$

This gives a Bayes classifier provided that the classification is performed according to the following rule:

$$f \in \mathrm{Cl}_m \text{ if } \hat{\Pr}(f \in \mathrm{Cl}_m) = \max_{l = 1, \ldots, k} \hat{\Pr}(f \in \mathrm{Cl}_l).$$

Note that in the practical implementation of this Bayes classifier, it is expedient to take the logarithm of the likelihood function. For the logarithm, we have

$$\ln(\hat{\Pr}(f \in \mathrm{Cl}_l)) = \ln(\hat{p}_l) + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \ln(\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j)\}). \quad (3.1)$$

Now let us consider the relation between the thus-obtained Bayes classifier and the distance functions considered in the previous section. Let us introduce a somewhat more complicated system of features for sign-based representations:

$$\pi_f^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1, & f(\mathbf{x}_i) \ge f(\mathbf{x}_j), \\ 0, & f(\mathbf{x}_i) < f(\mathbf{x}_j); \end{cases}$$

$$\pi_f^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1, & f(\mathbf{x}_j) \ge f(\mathbf{x}_i), \\ 0, & f(\mathbf{x}_j) < f(\mathbf{x}_i). \end{cases}$$

The features $\pi_f^{(1)}(\mathbf{x}_i, \mathbf{x}_j)$ and $\pi_f^{(2)}(\mathbf{x}_i, \mathbf{x}_j)$ contain the same information as the feature $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$; however, they are more useful in the fine adjustment of the generalized distance function d'_μ. In this case, the function μ is defined on all subsets of the set $\alpha \times \{1, 2\}$. For the images $f, g \in \mathcal{F}$, we have

$$d'_\mu(f, g) = \mu(\{(\mathbf{x}_i, \mathbf{x}_j, k) | \pi_f^{(k)}(\mathbf{x}_i, \mathbf{x}_j) \ne \pi_g^{(k)}(\mathbf{x}_i, \mathbf{x}_j)\}).$$

Note that the features $\pi_f^{(1)}(\mathbf{x}_i, \mathbf{x}_j)$ and $\pi_f^{(2)}(\mathbf{x}_i, \mathbf{x}_j)$ are always dependent. This follows, in particular, from the fact that they cannot vanish simultaneously. If, however, we assume that the features $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$ are independent, we find that

$$d'_\mu(f, g) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \mu(\{(\mathbf{x}_i, \mathbf{x}_j, k) | \pi_f^{(k)}(\mathbf{x}_i, \mathbf{x}_j) \ne \pi_g^{(k)}(\mathbf{x}_i, \mathbf{x}_j)\}).$$

It is seen that the distance function d'_μ makes it possible to assign less informativeness to a case when $f(\mathbf{x}_i) = f(\mathbf{x}_j)$ and $g(x_i) < g(\mathbf{x}_j)$, as compared to the case in which $f(\mathbf{x}_i) > f(\mathbf{x}_j)$ and $g(\mathbf{x}_i) < g(\mathbf{x}_i)$; these cases have the same informativeness with respect to the distance function d_μ.

Now, for any class Cl_l, $l = 1, \ldots, k$, we consider a etalon sign-based representation of $\mathcal{E}_l = (\pi_l(\mathbf{x}_i, \mathbf{x}_j))_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha}$ satisfying the condition

$$\pi_l(\mathbf{x}_i, \mathbf{x}_j) = \arg \max_{c \in \{-1, 0, 1\}} \hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c\}. \quad (3.2)$$

We also introduce the features

$$\pi_l^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0, & \pi_l(\mathbf{x}_i, \mathbf{x}_j) = 1, \\ 1, & \pi_l(\mathbf{x}_i, \mathbf{x}_j) \neq 1; \end{cases}$$
$$\pi_l^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0, & \pi_l(\mathbf{x}_i, \mathbf{x}_j) = -1, \\ 1, & \pi_l(\mathbf{x}_i, \mathbf{x}_j) \neq -1. \end{cases} \quad (3.3)$$

Then, assuming that features $\pi_f(\mathbf{x}_i, \mathbf{x}_j)$, $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$ are independent, we have

$$d'_\mu(f, \mathscr{E}_l) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \mu(\{(\mathbf{x}_i, \mathbf{x}_j, k) | \pi_f^{(k)}(\mathbf{x}_i, \mathbf{x}_j) \neq \pi_l^{(k)}(\mathbf{x}_i, \mathbf{x}_j)\}). \quad (3.4)$$

Proposition 2. *Classification by the rule*

$$f \in \mathrm{Cl}_m \text{ if } d'_{\mu_m}(f, \mathscr{E}_m) + \varepsilon_m = \min_{l = 1, \ldots, k} (d'_{\mu_l}(f, \mathscr{E}_l) + \varepsilon_l), \quad (3.5)$$

where $d'_{\mu_m}(f, \mathscr{E}_n)$ *is given by* (3.4) *for* $n = l, m$ *and* $\mu = \mu_m, \mu_l$, *and also*

$$\mu_n(\{(\mathbf{x}_i, \mathbf{x}_j, k) | \pi_f^{(k)}(\mathbf{x}_i, \mathbf{x}_j) \neq \pi_n^{(k)}(\mathbf{x}_i, \mathbf{x}_j)\}) = \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_n(\mathbf{x}_i, \mathbf{x}_j)\}) - \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_n}(\mathbf{x}_i, \mathbf{x}_j)\}), \quad (3.6)$$

$$\varepsilon_n = -\ln(\hat{p}_n) - \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_n(\mathbf{x}_i, \mathbf{x}_j)\}), \quad (3.7)$$

is equivalent to the application of a Bayes classifier. Also, distance function (3.4) *is a metric on sign-based representations if, for all* $(\mathbf{x}_i, \mathbf{x}_j) \in \alpha$, *the following conditions are satisfied*:

$$\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = 0\} \le \max_{c \in \{-1, 1\}} \hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c\}; \quad (3.8)$$

$$\left|\ln(\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = 1\}) - \ln(\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = -1\})\right| \le 2((\max_{c \in \{-1, 1\}} (\ln(\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c\})) - \ln(\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = 0\}))). \quad (3.9)$$

Proof. Assume that classification rule (3.5) is satisfied for some $m \in \{1, \ldots, k\}$; i.e., $f \in \mathrm{Cl}_m$ in the given classifier if

$$d'_\mu(f, \mathscr{E}_m) - \ln(\hat{p}_m) \le d'_\mu(f, \mathscr{E}_l) - \ln(\hat{p}_l)$$

for all $l \in \{1, \ldots, k\}$. Substituting expressions (3.6) into (3.5), this gives

$$\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} (\ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_m(\mathbf{x}_i, \mathbf{x}_j)\}) - \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_m}(\mathbf{x}_i, \mathbf{x}_j)\})) + \varepsilon_m$$
$$\le \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} ((\ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_l(\mathbf{x}_i, \mathbf{x}_j)\})) - \ln((\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j)\})) + \varepsilon_l.)$$

Putting (3.7) into the last inequality, we see that

$$\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_m}(\mathbf{x}_i, \mathbf{x}_j)\}) + \ln(\hat{p}_m)$$
$$\ge \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha} \ln(\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j)|\mathrm{Cl}_l\}) + \ln(\hat{p}_l)$$

or $\hat{\Pr}(f \in \mathrm{Cl}_m) \ge \hat{\Pr}(f \in \mathrm{Cl}_l)$ for all $l \in \{1, \ldots, k\}$; i.e., classifier (3.5) is equivalent to a Bayes classifier. Due to conditions for selection of a reference sign-based representation (3.2), we have $\hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_m(\mathbf{x}_i, \mathbf{x}_j)\} \ge \hat{\Pr}\{\pi_f(\mathbf{x}_i, \mathbf{x}_j) = \pi_{f|\mathrm{Cl}_m}(\mathbf{x}_i, \mathbf{x}_j)\}$, whence it follows that the proximity measure (3.6) is nonnegative.

From (2.1), (3.3) and (3.4) it follows that if $\pi_f(\mathbf{x}_i, \mathbf{x}_j) = 0$, then the single-element set $\{(\mathbf{x}_i, \mathbf{x}_j, 1)\}$ or $\{(\mathbf{x}_i, \mathbf{x}_j, 2)\}$ is the argument of function (3.6). If $\pi_f(\mathbf{x}_i, \mathbf{x}_j) = -1$ or $\pi_f(\mathbf{x}_i, \mathbf{x}_j) = 1$, then the two-element set $\{(\mathbf{x}_i, \mathbf{x}_j, 1), (\mathbf{x}_i, \mathbf{x}_j, 2)\}$ is the argument of function (3.6). Hence, (3.8) implies the condition $\mu(\{(\mathbf{x}_i, \mathbf{x}_j, k)\}) \le \mu(\{(\mathbf{x}_i, \mathbf{x}_j, 1), (\mathbf{x}_i, \mathbf{x}_j, 2)\})$ $k = 1, 2$, and (3.9) implies the condition $\mu(\{(\mathbf{x}_i, \mathbf{x}_j, 1), (\mathbf{x}_i, \mathbf{x}_j, 2)\}) \le \mu(\{(\mathbf{x}_i, \mathbf{x}_j, 1)\}) + \mu(\{(\mathbf{x}_i, \mathbf{x}_j, 2)\})$. This being so, by Proposition 2 we see that d'_μ is a metric on sign-based representations.

Remark 3. *We set* $\hat{\Pr}\{\pi_{f|\mathrm{Cl}_l}(\mathbf{x}_i, \mathbf{x}_j) = c\} = p_c$, $c \in \{-1, 0, 1\}$ *and assume, for definiteness that* $p_1 = \max_{c \in \{-1, 0, 1\}} p_c$. *Then condition (3.9) of Proposition 2 is equivalent to the condition* $p_0^2 \le p_1 p_{-1}$. *Hence, due to the symmetry of the last expression with respect to* p_1 *and* p_{-1}, *it follows that condition (3.9) implies condition (3.8). Note also that condition (3.9) appears quite adequate for real data provided that the quantization step of images is significantly smaller than the noise level, i.e., when* $p_0 \le \min_{c \in \{-1, 0, 1\}} p_c$.

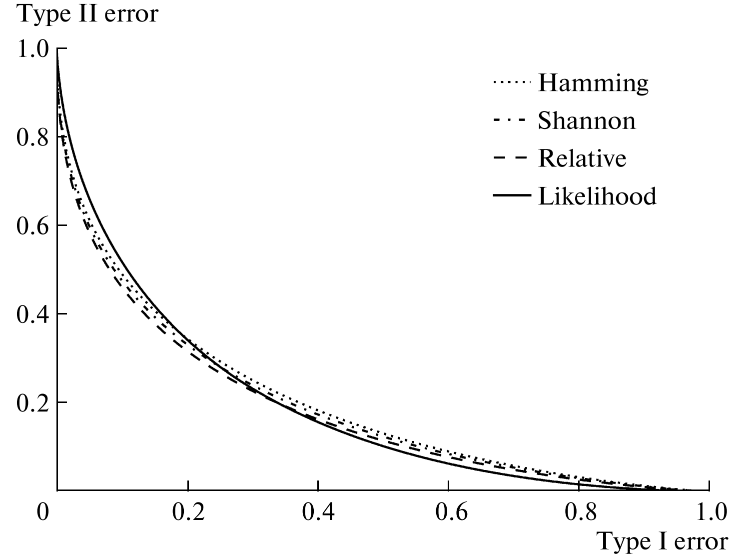

Fig. 1. Statistical estimates of errors of first and second kind for classifiers based on Hamming metric (Hamming), Shannon entropy (Shannon), relative entropy (Relative), and likelihood function (Likelihood).

4. CLASSIFICATION OF SIGN-BASED REPRESENTATIONS ON THE BASIS OF DISTANCE FUNCTIONS

The Bayes classifier, which was constructed in the previous section, is capable of ensuring a fairly general approach to classifying sign-based representations based on distance functions. This approach is as follows. For classes Cl_l, $l = 1, \ldots, k$, it is required to determine the etalon sign-based representations $\mathcal{E}_l$ in accordance with (3.2), as well as distance function d_l. Then the classification rule is given by

$$f \in \mathrm{Cl}_m \text{ if } d_m(f, \mathcal{E}_m) + \varepsilon_m = \min_{l = 1, \ldots, k} (d_l(f, \mathcal{E}_l) + \varepsilon_l); \quad (4.1)$$

note that the numerical parameters ε_l, $l = 1, \ldots, k$ in the last expression make it possible to control errors of first and second kind. Note that simplifications are possible in classification rule (4.1). In particular, we can set $\varepsilon_l = 0$, $l = 1, \ldots, k$, or assume that we are using the same distance function for each class; i.e., $d_l = d$, $l = 1, \ldots, k$. For example, as distance function d we can consider the metric based on the Shannon entropy (see Remark 2), which can be constructed either from system of features $\{\pi_f(\mathbf{x}_i, \mathbf{x}_j)\}_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha}$ or $\{\pi_f^{(k)}(\mathbf{x}_i, \mathbf{x}_j)\}_{(\mathbf{x}_i, \mathbf{x}_j) \in \alpha, k \in \{1, 2\}}$.

There are many ways to specify etalon sets of features. In particular, they can be chosen in the same way as for the Bayes classifier based on the chosen metric. It is noteworthy that in general the etalon sets of features that correspond to a Bayes classifier cannot be referred to any image; i.e., the set of images that corresponds to the reference sign-based representation may be empty. This, however, is not critical in recognition, since such a choice of a reference sign-based representation is in line with the optimal statistical test.

5. ESTIMATING CLASSIFICATION ERRORS OF SIGN-BASED REPRESENTATION OF IMAGES

Let us consider the binary problem of classifying sign-based representations of images and examine the behavior of classification errors of the first and second kind depending on the classifiers used. The following formulation of the problem is assumed.

Let $\mathcal{E}_0$ be the etalon image of the class Cl_0. An image f is referred to class Cl_0 if $d_\mu(f, \mathcal{E}_0) < d_0$; otherwise f lies in the alternative class Cl_1. Under real circumstances, an image f to be classified can be regarded as a realization of a multivariate random variable; hence, the sign-based representation τ of an image f is also a multivariate random variable. So, the value of the distance function $\xi = d_\mu(f, \mathcal{E}_0)$ can also be treated as a random variable.

In order to assess the separating capacity of the classifiers constructed above, we statistically estimate classification errors of the first and second kind. To illustrate this, we consider a face detection problem. We assume given a sample of images $X = X_F \cup X_N$, $X_F \cap X_N = \varnothing$, where X_F is a sample of facial images and X_N is a sample of arbitrary images not containing faces. From sample X it is necessary to estimate the probabilities $\alpha = \Pr\{\xi < d_0 | f \in \mathrm{Cl}_1\}$ and $\beta = \Pr\{\xi \geq d_0 | f \in \mathrm{Cl}_0\}$, which correspond to errors of the first and second kinds. When performing numerical experiments, we used the CBCL Face database[2] as sample X, the training subset of images was used to build classifiers, and the test subset [14] was used to estimate the separating capacity.

[2] http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html

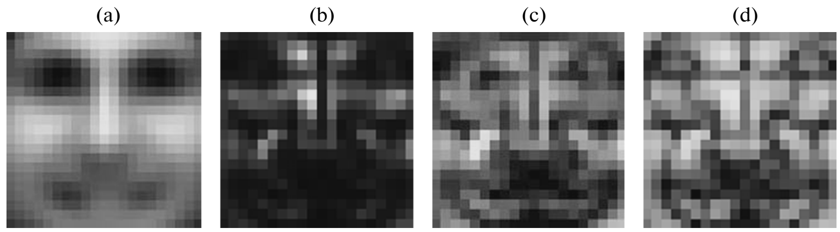

Fig. 2. (à) Image of reference face; (b)–(d) range maps of informativeness measure of features of sign-based representation based on: (b) Shannon entropy, (c) relative entropy, (d) likelihood function. The brighter the point, the higher its informativeness.

In order to estimate the probabilities $\hat{\alpha}$ and $\hat{\beta}$, we use the parametrical approach to estimate the density function of the distribution of random variable ξ; the parametrical approach consists in estimating parameters of some distribution from a sample. Since the value of ξ is formed as the sum of a large number of random variables (see (2.3)– (2.5), (3.6)), we assume that they are normally distributed.

From the plots in Fig. 1, one can see that the smallest classification errors correspond to classifiers on the basis of the relative entropy and the likelihood function. The metric based on the Shannon entropy and Hamming metric allows somewhat larger classification errors. This is not unexpected, because the distance functions based on the relative entropy take into account the distribution of probabilities of values of features for both classes Cl_0 and Cl_1, whereas the Hamming metric and the metric based on the Shannon entropy take into account only the distribution of probability of values of features of the class Cl_0. Figure 2 shows an image of the reference face and the normalized range maps for the information measure of features of sign-based representation, as estimated from the training sample. It is seen from the figure that the most informative face areas are the eye area, bridge of the nose, and cheekbones, which is in accordance with intuition.

6. CONCLUSIONS

The paper proposes a new way to introduce distance functions on sign-based representations described by a set of features based on classical functionals of information theory. The constructed metrics is equivalent to the Hamming metrics in the simplest case when it is assumed that features determining the sign-based representation are independent and have the same informativeness. However, more statistically optimal distance functions are obtained if the informativeness of features is measured using the Shannon entropy or Kullback–Leibler distance. In the latter case, the distance function is capable of assigning larger informativeness to features from which the given class is different from the other; this results in reduced classification errors. The construction of a classifier of sign-based representations based on likelihood functions is considered. We show that this classification is also based on a special type of distance functions. In the concluding part of the paper, we consider application of the constructed distance functions to the problem of classifying images in two classes: "face" and "nonface." We have obtained statistical estimates of errors of the first and second kinds to show that the classifier based on the relative entropy is the best among the constructed classifiers. Classification based on likelihood functions produces nearly the same results. This is followed by the classifier based on the metric with the Shannon entropy and the classifier based on the Hamming metric.

The proposed approach to constructing distance functions is quite general and can be applied to construct distance functions between images described by an arbitrary system of symbolic features.

This paper is a revised and expanded version of a paper presented at the scientific conference IOI-8 (Intellectualization of Information Processing) [1].

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 08-07-00129, 10-07-00135, and 10-07-00478.

REFERENCES

1. A. G. Bronevich and A. V. Goncharov, "The Way to Classify Images Character Representation," in *Proc. 8th Int. Conf. Intellectual Processing for Information. Cyprus, Pathos, Oct. 17–24, 2010* (Maks Press, Moscow, 2010), pp. 321–324 [in Russian].
2. A. Goncharov and V. Gubarev, "Comparison of High-Level and Low-Level Face Recognition Methods," in *Proc. 9th Conf. on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-9-2008)* (Nizhni Novgorod, 2008), vol. 1, pp. 178–181.
3. A. V. Goncharov and V. V. Gubarev, "The Way to Detect Faces Specific Features at Digital Images by Means of Character Representation," in *Proc. All-Russian Conf. MMPO-14* (Suzdal, 2009), pp. 325–328.
4. A. Goncharov and A. Melnichenko, "Pseudometric Approach to Content Based Image Retrieval and near Duplicates Detection," in *Proc. Russian Seminar on Information Search Methods Estimation. Works of ROMIP: Russian Information Retrieval Evaluation Seminar* (2007-2008), pp. 120–134.
5. A. G. Bronevich and A. V. Goncharov, "Images Character Representation and Its Self-Descriptiveness," in *Proc. All-Russian Conf. MMPO-14* (Suzdal, 2009), pp. 309–312.

6. A. G. Bronevich and A. V. Goncharov, "Axiomatic Approach to Measure Self-Descriptiveness of Images Character Representation," Izv. Akad. Nauk. Teor. Sist. Upr., No. 6, 206–218 (2010).
7. A. N. Karkishchenko and A. V. Goncharov, "Images Self-Descriptiveness Stability," Avtomat. Telemekh., No. 9, 57–69 (2010).
8. E. Jacobs, A. Finkelstein, and D. Salesin, "Fast Multiresolution Image Querying," in *Proc. Annu. Conf. Computer Graphics* (Los Angeles, 1995), pp. 277–286.
9. O. Jesorsky, J. K. Klaus, and W. F. Robert, "Robust face detection using the Hausdorff Distance," Lecture Notes Comput. Sci. Audio- Video-Based Peron Authentication (AVBPA) **2091**, 90–95 (2001).
10. B. G. Litvak, *Expert Information: Ways to Get and Analyze* (Radio i svyaz', Moscow, 1982) [in Russian].
11. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley-Intersci., 2001).
12. E. Excolano, P. Suau, and B. Bonev, *Information Theory in Computer Vision and Pattern Recognition* (Springer-Verlag, 2009).
13. D. Dennebery, *Non-Additive Measure and Integral* (Springer, 1994).
14. B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, "Component-Based Face Recognition with 3D Morphable Model," in *Proc. 1st IEEE Workshop on Face Processing in Video* (Washington, 2004).

Translated by A. Alimov

Andrey G. Bronevich. Born 1966. Graduated from Taganrog State University of Radio Engineering 1988. Received candidate's degree 1994 and doctoral degree 2004. Professor at National Research University Higher School of Economics. Scientific interests: nonadditive measures, theory of imprecise probabilities, decision theory under risk and uncertainty, pattern recognition, image processing. Member of Society for Imprecise Probability: Theories and Applications. Member of editorial board of *International Journal of General Systems*. Author of more than 90 papers.

Alexander V. Goncharov. Born 1984. Graduated from Taganrog State University of Radio Engineering 2006. Received candidate's degree n 2010. Chief research and development officer at CVisionLab, LCC. Scientific interests: pattern recognition, machine learning, content-based image retrieval.