Telecommunications and Radio Engineering, 73 (18):1661-1670 (2014)

QUALITY CRITERIA FOR MULTIDIMENSIONAL OBJECT RECOGNITION BASED UPON DISTANCE MATRICES

A.V. Gorokhovatsky, V.A. Gorokhovatsky♦, A.N. Vlasenko, & N.V. Vlasenko

Kharkiv National University of Radio Engineering and Electronics, 14, Lenin Ave, Kharkiv, 61166, Ukraine

♦Address all correspondence to V.A. Gorokhovatsky E-mail: gorohovatsky-v@rambler.ru

The problems of evaluation of the image recognition quality in terms of applying structural methods are addressed. The similarity matrice between the descriptions of objects for a finite base of images as the basic source of criterion formation is suggested. An approach to optimizing the proposed criteria is discussed. The results of experiments on calculating the probability of correct recognition are examined.

KEY WORDS*: multidimensional object, structural image description, similarity matrix, recognition correctness criterion, optimization of a threshold for elements equivalence, noise immunity*

1. INTRODUCTION

The problem on recognition of multidimensional objects, in particular, images involve the following interrelated critical factors [1-3]. Specifically, they fear on a certain method to be applied, space of attributes and the base of images in the scope of which the recognition operation is being performed. We can state with assurance that the results evaluated through some recognition quality criterion is largely governed by the consistency of these key factors. In other words, it helps find ant to whet extent the method to be used and the space of attributes were aptly chosen for a specific set of recognizable objects. We believe that as far as the classification problem (with a finite number S of object classes) the quality can be estimated in terms of calculating an experimental $S \times S$ confusion matrix $< \beta\left(C_i, C_k\right) >$ [1], which for each pair of classes ($i \neq k,\ i,k = \overline{1,s}$) indicates how many objects of class C_i are erroneously attributed to class C_k, i.e., it actually accounts for the number of erroneous decisions. The matrix

$<\beta\left(C_i,C_k\right)>$ is usually constructed on the basis of the data gleaned from the tests of learning or control samples of objects. Meanwhile the confusion matrix contains the diagonal elements $<\beta\left(C_i,C_i\right)>$ that specify the number of correct responses. In terms of confusion matrices one can evaluate the probabilities of errors and a correct classification by dividing its value into a total number of sampling elements. In general the confusion matrix serves to find out whether it is in fact necessary and possible to upgrade the classification quality. For instance, the database contains two objects that bear a close resemblance from the standpoint of the method being applied, then in order to minimize the error level, one of these images (base modification) needs to eliminated or the space of attributes and/or the classification method should be altered as a whole.

An interclass distance matrix being well-used in clustering problems, is lose conceptually to notion, albeit differing from the confusion matrix. This matrix contains the values of $<\rho\left(C_i,C_k\right)>$ distances (or the measures of similarity) between the pair of classes. It is intuitively clear that the distances $\rho\left(C_i,C_k\right)$ should be as great as possible at $i \neq k$. This signifies that different classes are materially distingue shamble, and these distances should be as short as possible at $i = k$, which is indicative of the objects being close to each other in one and the same class. According to the interpretation of the statistical decision theory [3] the value of both $<\rho\left(C_i,C_k\right)>$ and $<\beta\left(C_i,C_i\right)>$ account for the errors of the first and second find, which makes it possible to calculate the probability of correct classification.

The matrix of pair wise distances between the elements of sets of multidimensional objects was found to be also feasible in multidimensional data visualization problems [4] aiming to evaluate the quality of representing the sets of objects in a synthesized space. Considering that herein the key factor is the level of data structure variability following the mapping into a new space, this particular space is constructed through minimization of a functional that typifies the distance matrix deviations for these spaces. The visualization error is evaluated as follows:

$$v = \sqrt{\frac{2}{s\left(s-1\right)}\sum_{i=1}^{s-1}\sum_{j=i+1}^{s}\left(\rho\left(X^i,X^j\right)/V - \rho\left(\tilde{X}^i,\tilde{X}^j\right)/\tilde{V}\right)^2}\ , \qquad (1)$$

where s is the number of objects, $X^i,\tilde{X}^i$ is the initial transformation of space, $V = \max\limits_{\substack{i=\overline{1,1-s}\\ j=\overline{i+1,s}}} \rho\left(X^i,X^j\right)$, $\tilde{V} = \max\limits_{\substack{i=\overline{1,1-s}\\ j=\overline{i+1,s}}} \rho\left(\tilde{X}^i,\tilde{X}^j\right)$ are the normalizing factors reducing the matrix values to interval [0,1].

As seen from the foregoing, the error as given in (1) is determined by the averaged normalized Euclidean distance between the distance matrix elements in two data-representation spaces $X^i,\tilde{X}^i$. Expression (1) takes account of the distances between the objects of different classes. This indicates that in view of the distance matrix symmetry the different objects are erroneously close to each other.

As will be apparent from the concepts that have been analyzed and generalized, one can state that the confusion and interclass distance matrices represent a common principle in evaluating the closeness/distinction of objects in a certain space of their description. In terms of calculating the definite matrix characteristics the above principle can be used as the basis for constructing the numerical criterion which is indicative of the objects classification quality, as the specific procedure is utilized in a preassigned space object descriptions drawn from the database.

2. DISTANCE MATRIX-BASED CONSTRUCTION AND VALIDATION OF THE CRITERION

Let $O=\{O_k\}$ be the set (base) of objects; $Z=\{Z_k\}$ be the set of their descriptions (space of attributes; $E=\{E_k\}_{k=1}^{N}, E\subseteq O$ be the finite set of standard objects specifying the number N of classes in the base; M be the classification method. Let the classification be assigned as mapping $M:O\to Z\to E$, thereby determining the number of class $k=\overline{1,N}$ of an object. In this case the classification quality criterion Υ is the function $\Upsilon=\Upsilon(M,O,Z,E)$ (see Fig. 1).

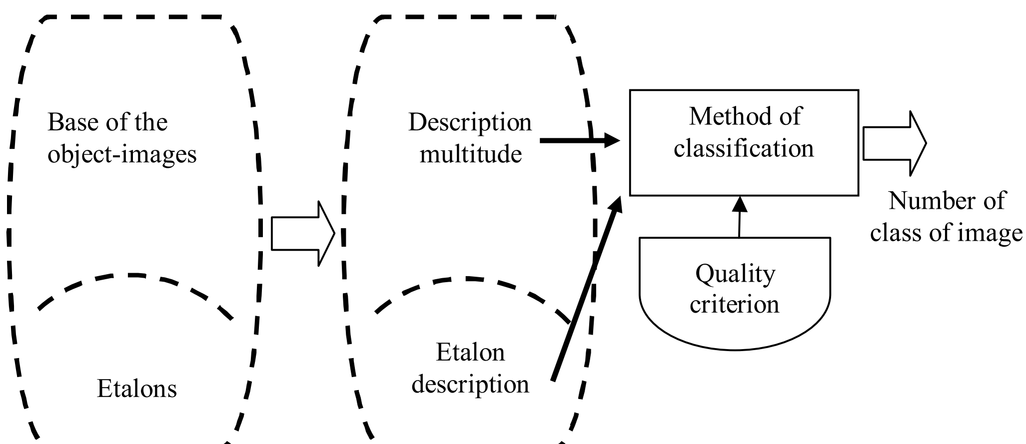

FIG. 1: The schematic diagram illustrating the use of the classification quality criterion

Variations in any of parameters M,O,Z,E affects the quantity of Υ. Specifically, as the space of attributes Z is being transformed, the changes in the property of the whole classification system is brought about, thereby modifying Υ [5]. The construction of a new space is generally dictated by the need to attain the more optimal recognition characteristics which are more frequently representative of the snap-action and noise immunity criteria. In the meantime note that matrix $\{\rho(C_i, C_k)\}_{i,k=1}^{N}$ is thereupon calculated both in the initial and in transformed spaces of attributes for one and the same sample of objects to be classified.

Now let us construct the quality criterion Υ in terms of calculating the distance matrix $R = \{\rho(C_i, C_k)\}_{i,k=1}^{N}$, which, apart from parameters M, O, Z, E discussed above is therewith the function of objects sampling from which it was obtained. On frequent occasions, matrix R is constructed for a set E of reference objects and points to the level of standard distinction in the space of attributes for the base of images. Matrix R and the sample for which it is obtained is associated with geometrical transformations of input images, the level and type of interferences and other factors [9].

Now let us make the problem of constructing the criterion $\Upsilon[R]$ more specific to evaluate the classification quality through the analytical treatment of the structural description of an image as a set of characteristic features [6]. We will make a decision on the objects class according to a maximum of its similarity on a set of descriptions from N reference objects. When classifying the elements of structural descriptions of objects using the procedure of element voting, we make use (in lien of the distances between descriptions) of the description similarity as the number $h(C_i, C_k)$ of votes given by the description of object C_k for reference C_i [6]. The similarity $h(C_i, C_k)$ based on the voting function does not have symmetry as compared to the metric, and as regards this similarity, in the general case, the triangle inequality is not satisfied [2]. However, in this instance, function h is calculated in a simpler and more practicable fashion.

Therefore we will take a look at the integer matrix of votes $H = \{h_{ik}\} = \{h(C_i, C_k)\}_{i,k=1}^{N}$ along with the distance matrix. Table 1 shows a practical example of matrix H for $N = 12$ reference images of the animals base, which were described as a set of characteristic features SURF when the set of rotation-transformed standards was analyzed. In this particular instance the standards have quantitively different make-up of description components. It is seen that the significant similarity of the standard No 5 with other standards is inherent in this base, because the interclass similarity is as high as 26/62 = 42%. If a diagonal element exceeds others in a row (column), then the classification is held to be correct, because a decision is taken according to a maximum number of votes. In the general case matrix H is not symmetrical and unification of its calculation is determined by the standardization of the construction procedure. For example, the number of votes h_{ik} is calculated to one side as the quantity of elements of an i-th reference object, which were found to be similar in an analyzed object with number k. In this situation the standard acts as if it "draws" similar elements of recognizable description.

TABLE 1: An example of the voting-based similarity matrix

67	2	1	0	2	0	0	5	2	2	0	2
	84	1	0	5	5	6	9	0	8	1	3
		27	0	0	1	0	2	3	1	1	0
			159	1	9	5	3	0	0	6	0
				62	2	1	25	0	26	1	12
					92	11	6	1	0	14	0
						91	5	1	1	6	0
							56	0	8	2	9
								54	0	0	0
									79	2	13
										76	1
											36

In terms of improving the classification quality the values of similarity h_{ik} should be as large as possible at $i = k$ (diagonal). This represents the closeness of objects from one class in a constructed space of attributes. Also these value should be as low as possible at $i \neq k$, which indicates that the objects of different classes are appreciably distinguishable from each other. Since these standards have a priori a different number of votes (g_i is the number of elements in the description of the standard with number i), it makes sense to previously normalize the diagonal elements of matrix H and they to pass to relative values $h_{ii}^* = h_{ii} / g_i$, which are equivalent to the proportion of obtained votes for "their" standard with respect to their maximum number g_i. As far as the elements h_{ik} at $i \neq k$ are concerned, the estimate given as $h_{ik}^* = h_{ik} / h_{ii}$ is meaningful. It indicates to what extent the values of nondiagonal elements are in excess of a diagonal element. It is apparent that if the values of h_{ik}^* is large, i.e., it exceeds, say, $h_{ik}^* \geq 0.5$, then it is highly probable that there is a significant similarity of the description of the base with numbers i, k . This immediately leads to a decreased probability of correct recognition (especially, under interference conditions) due to the initially high likelihood of the event of a false alarm.

As a result, our reasoning behind the foregoing facts necessitates the joint analysis of two inconsistent criteria A_1, A_2 that have an immediate effect upon the degree of probability of correct recognition using the given method in a particular image base. These criteria are formed around matrix H and given as functions $A_1\left(h_{ii}^*\right)$ and $A_2\left(h_{ik}^*\right)$, the probability of correct being directly proportional to variation in h_{ii}^* and inversely proportional to value of h_{ik}^*. In this context the aggregated quality criterion Υ can be specified as the dependence $\Upsilon = \Upsilon\left[A_1\left(h_{ii}^*\right), A_2\left(h_{ik}^*\right)\right]$ upon two partial criteria.

3. THE TWO-CRITERIA APPREACH TO OPTIMIZING THE THRESHOLD PARAMETER

Now let us elucidate the formation of aggregated criterion Υ in the problem on the optimal selection of the threshold for the equivalence of description elements in comparing the feature representation as the sets of characteristic image features [7]. In examining the properties of descriptions as the sets of detector-vectors in terms of the SURF method, it was found that the quality of image classification is largely governed by the threshold value of δ_z that serves to assign the equivalence of description elements (SURF vectors). The decision on equivalence is appropriate to the predicate truth $L\left[\rho\left(z_1,z_2\right),\delta_z\right]$. Once the distance between elements z_1,z_2 is shorter than the threshold, the elements z_1,z_2 are taken to be equivalent, i.e.,

$$L\left[\rho\left(z_1,z_2\right),\delta_z\right]=\begin{cases}1, & \rho\left(z_1,z_2\right)\le\delta_{z,}\\ 0, & \text{otherwise}.\end{cases}$$

Based upon δ_z, the similarity value between descriptions is calculated as the number of votes given by the elements of the etalon for a recognizable object. As regards the domestic animal image base in use, the range of experimental threshold values of δ_z amounted to an interval [0.08; 0.8]. At $\delta_z=0.8$ and more all the conformities of points of a transformed image are specified in the process of comparison. This clearly demonstrate that the detector invariance is fully satisfied, i.e., it is extended to all point. This invariance is kept at the threshold value of $\delta_z<0.8$ solely for a certain part of description points. Here considering $\delta_z=0.08$ and under, the number of invariant points does not go beyond 5% of the total volume of description of the standard, which is likely to result in some difficulties over the discrimination of objects in the space of attributes. Now let us focus on selecting an optimal threshold $\delta_z\in\left[0.08;\ 0.8\right]$, allowing one to make a reliable estimate of similarity while ensuring the objects of different classes remain distinguishable.

The examination of vote matrices H suggested that, as the threshold value of δ_z increases over the indicated range, the elements h_{ij} get scaled up. This is indicative of the growing similarity between different standards. Thus, under the present circumstances it is necessary that the values of two partial criteria be optimized in keeping with matrix H: herein we are to meet the requirement that the maximum largest values of diagonal elements h_{ii} be retained with simultaneously reduced values of remaining elements h_{ij}.

Now we would like to formulate our reasoning by way of criteria $A_1\left(\delta_z\right)$ that need to be maximized and criteria $A_2\left(\delta_z\right)$ that also need to be minimized, thereby arriving at the following relations:

$$\max A_1(\delta_z) = \frac{1}{N}\sum_{i=1}^{N} h_{ii}^{*}, \tag{2}$$

$$\min A_2(\delta_z) = \frac{1}{N-1}\sum_{i=1}^{N-1} 1\left(\max_{i=1,\dots N-1} h_{ik}^{*} \geq 0.5\right), \tag{3}$$

where N is the number of etalons of the base, A_1 is the averaged ratio between the over-number of standards of matrix diagonal elements and their maximum values; A_2 is the fraction of the number of matrix H rows in which the maximum closest to the diagonal value is greater than the half of diagonal element h_{ii} whereas the function $1(.)$ for the matrix row in (3) is defined as follows

$$1\left(\max_{i=1,\dots N-1} h_{ik}^{*} \geq 0.5\right) = \begin{cases} 1, & \max_{i=1,\dots N-1} h_{ik}^{*} \geq 0.5, \\ 0, & \max_{i=1,\dots N-1} h_{ik}^{*} < 0.5. \end{cases} \tag{4}$$

For simplicity sake, the similarity matrix symmetry is assumed in relation (3). To be more specific as regards the formulation of (4), exceeding of a nondiagonal element over a diagonal one by more than 0.5 is taken to be meaningful. Basically, this magnitude is an individual parameter that can be adapted to the image base.

Criterion A_1 is indicative of the need to retain the greatest possible number of votes given by the description elements for their class. Criterion A_2 serves to minimize the similarity between different standards of the base in the space of attributes. Thus, the reliable estimation of the description similarity can be represented as the two-criteria optimization problem with in consisting partial criteria (2), (3). Note that using the averaging operations in (2) and counting of meaningful elements is only one of the possible alternatives of analyzing the set of matrix elements that generate a diversity of possible criteria.

By scrutinizing the existing methods of multicriteria optimization [8] we are led to conclude that in order to find an optimal solution it would be judicious to make use of the criterion convolution ensuring that the utility function is optimized. To this end we shall reduce the two-criteria problem to aggregated criterion A by convolving criteria A_2, A_1 into a single complex one given as

$$A = \min(A_2 - A_1) \tag{5}$$

considering that the values of both partial criteria are within an interval [-1, 1].

4. SIMULATED RESULTS

To determine an optimal threshold value according to criterion (5) we have calculated the interclass similarity matrix by voting for the base 12 of domestic animal images using the detector of characteristic features [5,6]. Ideally, when there are no interference, the number of votes given by the reference standard offers a spread of points ranging between 100 and 200 within the base. This signifies that the number of points extracted by the detectors for different standards is rather uneven. The base under study exhibited some singularity: two etalons (No1 and No5) show a close similarity. Therefore, even if there are no interferences, 20 similar points are specified to describe them.

All etalons were further subjected to geometrical modifications (rotation by 20^0) and once again the interclass distance matrix was calculated for each threshold value on the interval [0.08; 0.8], namely, for discrete values $\delta_z = \{0.08; 0.16; 0.32;\ 0.4;\ 0.64;\ 0.8\}$.

As the threshold value δ_z grows, the diagonal values of vote matrix decreases. At the same time the number of votes given for "strange" standards tends to increase. The analysis of the experimental data that were collected suggested that in terms of maximizing the votes (the number of similar points) the optimal threshold is $\delta_z = 0.8$ (maximum of criterion A_1). As far as minimization of A_2 is concerned, the optimum value is 0.08. An optimal value of integrated criterion $A = \min(A_2 - A_1)$ is attained with the threshold 0.32 (it is equal to -0.6). In this context the total reduction of votes is on the order of 25% whereas in the general case $h^*_{ii} = 0.3$ and $h^*_{ii} = 0.6$ far the standards No1 and No5 that bear the closest similarity. Figure 2 presents the graphical interpretation of the threshold-dependent A, A_1, A_2 criterion values.

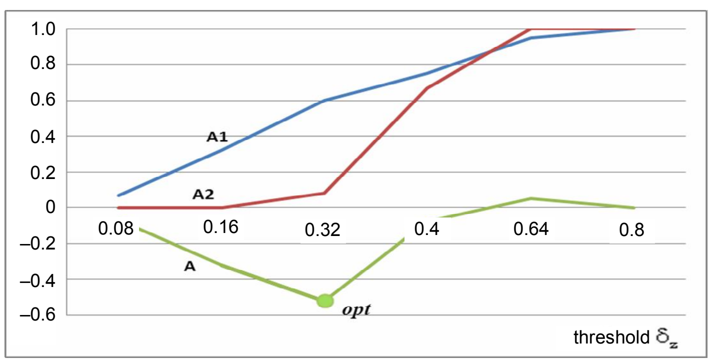

FIG. 2: The δ_z -dependent criteria A, A_1, A_2

We have made an experimental estimate of the interference-free recognition quality for a chosen animal image base using the transformation of rotation through a set of angles between 20 and 100^0 with respect to the standard at different thresholds of δ_z. The findings of the study suggested that the error-free recognition is seen to occur practically over the entire range of variations in δ_z. However, at values close to 0.08

one can see an insignificant number of characteristic features from which a decision on a particular class is taken. Yet at values close to 0.8 a slight difference between the peak values of the number of votes and the nearest to-it local maximum. In both instances we have to deal with an unsteady decision, which leads to a decrease in the degree of certainty, particularly, in the solid base of images as well as in the situation that are similar in terms of describing standards. The significance of this factor is directly governed by the standards held in the base.

The following series of experiments validate the efficiency of using the proposed criterion in an effort to increase the probability P of correct recognition under the additive interference conditions. The aim of the experiment we have conducted was to estimate the noise immunity level for the recognition technique based upon the similarity computation involving the optimal threshold chosen in keeping with a minimum of criterion A as given by (5). We have arrived at the probabilities of correct recognition (see Table 1) for the additive noise-polluted images from the domestic animal image base (Fig. 3, [7]) as function of the noise r.m.s. deviation σ at different thresholds δ_z and a fixed parameter of image rotation through 20^0.

FIG. 3: The examples of images taken from the domestic animal base

TABLE 2: The noise level-dependent recognition probability P at different thresholds

Noise level, σ	$\delta_z = 0.16$	$\delta_z = 0.32$	$\delta_z = 0.48$
38	0.99	1.0	0.97
48	0.87	0.99	0.87
57	0.69	0.96	0.76

As seen from Table 2, at an optimal value of $\delta_z = 0.32$ the magnitude P is substantially greater for all noise level and at $\sigma = 57$ is equal to 0.96 whereas for other thresholds it is appreciably lower and runs as high as 0.69 and 0.76. These ratios are indicative of efficient application of the optimal threshold in keeping with the criterion minimum (5). Moreover, we should like to place an emphasis upon the relatively high noise immunity level, because $\sigma = 57$ checks with the amplitude signal-noise ratio of 1.5 (image 250 x 220, mean brightness 75).

5. CONCLUSIONS

The studies that we have made and the results thus achieved have led us to get the quality criterion formalized in terms of the similarity matrix for a finite set of reference etalons from the image base. Besides, for all practical purposes, we were able to offer an opportunity to apply the formed criterion through the solution of the two-criteria optimization problem in an effort to make an optimal of the threshold for equivalence of description elements in establishing the similarity of visual objects in the process of recognition. The values of the proposed partial criteria and the aggregated criterion provide direct evidence of the efficiency of the recognition method suitable for applied problems of computer vision.

The experiments we have conducted suggested that the image recognition probability is substantially higher in deciding on an optimal threshold value under interference action than at other thresholds value.

What is lying behind the prospects for further studies is the need to generalize the proposed matrix criteria in case the partial criteria and image bases are arbitrarily selected.

REFERENCES

1. Manning, C.D., Raghavan, P., and Schutze, H., (2008), *Introduction to information retrieval*, Cambridge, University Press, - 528 p.
2. Ayvazyan, S.A., Buchstaber, V.M., Yenyukov, I.S., and Meshalkin, L.D., (1989), *Applied statistics: Classification and decreasing of dimensions*, Finance and statistics, Moscow: 607 p. (in Russian).
3. Duda, R.O., Hart, P.E., and Stork, D.G., (2000), *Pattern classification*, Willey, - 738 p.
4. Shklovetz, A.V. and Aksak, N.G., (2012), Determination of distances between the points Cohonen on a piecewise-smooth maps, *Bionika intellekta*, **1**(78):63-67 (in Russian).
5. Vlasenko, N.V. and Sytnik, O.V., (2013), Classification of video-objects in attribute space of the Walsh functions, *Telecommunications and Radio Engineering*, **72**(19):1777-1785.
6. Gorokhovatsky, V.A. and Putyatin, Ye.P., (2009), Image likelihood measures of the basis of the set of conformities, *Telecommunications and Radio Engineering*, **68**(9):763-778.
7. Vlasenko, N.V., (2013), Optimal multicriteria selection of threshold parameter in comparing the descriptions of objects, *Int. Conf. "Intellectual decision-making system and the problems of computer intellect" (ISDMCI)*, Kherson, pp.423-425.
8. Thomas, L.S., (2001), *Decision making with dependence and feedback: The analytic network process*, University of Pittsburgh. RWS Publications, -370 p.
9. Shapiro, L. and Stockman, G., (2001), *Computer vision*, Prentice Hall, - 625 p.