Proceedings of IAM, V.15, N.1, 2026, pp.140-150
https://doi.org/10.30546/2225-0530.15.01.2026.140

EMPIRICAL EVALUATION OF ADEQUACY CRITERIA FOR IMAGE RECOGNITION

O.A. Aliyev[1]

[1]Baku State University, Baku, Azerbaijan
e-mail: orkhanali@bk.ru

Abstract. This paper presents an empirical investigation of adequacy criteria used for evaluating image recognition methods based on artificially generated families of images. The study aims to analyze the reproducibility, robustness, and interpretability of these evaluation criteria within a strictly controlled computational experiment involving predefined distortion parameters and systematically generated image transformations. Artificial image families are constructed through sector-based partitioning of national emblem images, followed by linearization and feature extraction using the continuous wavelet transform. This procedure preserves the essential spatial and structural properties of the images while enabling the construction of a consistent feature representation suitable for quantitative analysis. Each image is subsequently transformed into a one-dimensional signal representation, allowing the application of signal analysis techniques and enabling systematic computation of similarity measures in the wavelet feature space. Euclidean distances between feature vectors are used to construct distance matrices that form the basis for quantitative analysis of recognition performance, algorithm stability, and localized structural correspondence between images. The use of artificial image families makes it possible to investigate the behavior of recognition methods under controlled structural distortions and to analyze the sensitivity of evaluation criteria to systematic variations in input data. Experimental results obtained for several artificial image families demonstrate that the proposed evaluation framework consistently identifies structurally coherent correspondences, adequately reflects the influence of controlled distortions, and reveals limitations of conventional global similarity measures that rely solely on holistic comparison of images. The study confirms the applicability of the considered adequacy criteria to a new class of structurally complex images and demonstrates their reproducibility across different experimental configurations. The results also indicate that the proposed evaluation framework can serve as a reliable and interpretable tool for systematic and objective assessment of image recognition methods in controlled experimental settings.

Keywords: Image recognition, artificial image families, continuous wavelet transform, sector-based image analysis, signal linearization, empirical validation.

AMS Subject Classification: 68T10, 94A08.

1. Introduction

In image recognition tasks, one of the key and still unresolved challenges is the objective assessment of the adequacy of the algorithms employed. Despite the rapid development of machine learning and computer vision techniques [2, 4], the problem of obtaining a reliable and reproducible evaluation of algorithm behavior under varying conditions remains insufficiently formalized. In most studies, the performance of recognition methods is assessed using integral metrics such as accuracy, sensitivity, and recall. Although these indicators characterize the final classification results, they do not always provide insight into the internal behavior of algorithms when input data undergo systematic transformations [12, 6, 1]. In particular, such metrics typically reflect only the final performance of a method and provide limited information about the stability of algorithms under controlled geometric or structural modifications of input images. In order to address this limitation, a set of formalized adequacy criteria for evaluating recognition methods was proposed in [11]. These criteria are based on the analysis of artificially generated families of signals and images and make it possible to investigate how distances between patterns evolve in feature space when reference images undergo sequential spatial displacements. The proposed criteria characterize several important aspects of algorithm behavior, including the uniformity of distance variation, sensitivity to perturbations, convergence dynamics, and overall stability of recognition procedures. This approach enables a transition from predominantly heuristic performance assessment toward a quantitatively justified analysis of recognition algorithms under controlled distortions.

Previous studies have demonstrated the practical applicability of these criteria in the analysis of recognition methods based on dynamic time warping, wavelet transform, and Fourier transform [9, 7, 3]. The obtained results confirmed that the proposed criteria can be used for the comparative evaluation of algorithms of different nature and computational principles. However, the question of their universali_This discrepancy indicates that the employed global comparison procedurety and robustness when applied to different feature extraction strategies, data representation schemes, and image classes remains open [5]. In particular, it is still unclear whether these criteria retain their informativeness when applied to images characterized by complex spatial structures and pronounced symmetry.

In the present work, an additional empirical verification of the adequacy criteria is performed using a new artificial family of images constructed from national emblem images. This class of objects was selected due to its high structural complexity, the presence of both central and axial symmetry, and the hierarchical organization of graphical elements. Such properties make national emblems a convenient model for studying the sensitivity of recognition algorithms to both local and global structural changes.

Unlike most existing studies that focus primarily on comparing specific recognition algorithms, the emphasis of the present work is placed on the reproducibility and robustness of the evaluation criteria themselves. Thus, the object of investigation is not the efficiency of a particular recognition method but the validity and universality of the evaluation framework used to analyze algorithm behavior. Continuous wavelet transform is employed to construct feature representations of images; however, within the context of this study it is treated solely as a data processing tool that allows the extraction of informative features rather than as a subject of comparative methodological analysis.

The experimental investigation is conducted using artificially generated image series obtained through controlled spatial displacements of reference patterns. Such an approach ensures strict regulation of experimental conditions, reproducibility of results, and the possibility of systematic analysis of adequacy criteria under predefined distortions. Particular attention is devoted to examining the behavior of the proposed criteria when varying the parameters used to construct artificial image series, as well as the structural properties of the images and the characteristics of the feature space.

The objective of this study is to perform an extended empirical validation of adequacy criteria for image recognition methods using an artificial family of national emblem images and continuous wavelet transform as a feature extraction tool [9]. The scientific novelty of the work lies in the systematic experimental verification of the universality and reproducibility of adequacy criteria on a new class of structurally complex images. In addition, the study identifies the conditions under which these criteria preserve their informativeness, interpretability, and stability when the parameters of image representation and feature extraction are varied [8].

2. Problem formulation

In the present study, the analysis of the adequacy of image recognition methods is based on a system of formalized evaluation criteria that have been previously proposed and experimentally validated in a number of studies. These criteria are designed to provide a quantitative characterization of the behavior of recognition algorithms when analyzing artificially generated families of signals and images. Their use makes it possible to study how recognition methods respond to controlled structural transformations of reference patterns and to assess the stability of similarity measures in feature space.

Let an artificial family of signals $S=\{s_1, s_2, \ldots, s_{Nk}\}$ be given, constructed by successive spatial displacements of a reference signal s_0 with a fixed step size h. Denote by $D_k(s_i, s_j)$ the Euclidean distance between signals s_i and s_j, computed using the k-th recognition method, where $k=1, 2, \ldots, K$ is the index of the method. For each signal s_i, the distance to the reference signal s_0 is defined as $d_i^{(k)} = D_k(s_i, s_0),\ \ i = 1, \ldots, N_s$. Based on this notation, the following adequacy evaluation criteria are introduced [6].

Criterion (C_1) (uniformity). The uniformity criterion reflects the requirement that distances from the reference signal increase monotonically and in a regular manner as the signals move away within the artificial family. Formally, a sequence of increments is introduced

$$\Delta_i^{(k)} = d_{i+1}^{(k)} - d_i^{(k)},\ \ i = 1, \ldots, N_s - 1. \tag{1}$$

The mean value of the increments is defined as

$$\bar{\Delta}^{(k)} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s - 1} \Delta_i^{(k)}. \tag{2}$$

The degree of uniformity is then estimated using the standard deviation

$$\sigma_{C_1}^{(k)} = \sqrt{\frac{1}{N_s - 1} \sum_{i=1}^{N_s - 1} \left(\Delta_i^{(k)} - \overline{\Delta}^{(k)} \right)^2} \ . \tag{3}$$

Smaller values of this measure correspond to higher adequacy of the recognition method with respect to criterion C_1.

Criterion (C_2) (symmetry). The symmetry criterion characterizes the consistency of distances to neighboring signals within the artificial family. For each internal element of the family, the following ratio is defined

$$r_i^{(k)} = \frac{D_k(s_i, s_{i+1})}{D_k(s_i, s_{i-1})}, \ i = 2, ..., N_s - 1 . \tag{4}$$

The average value of the ratio is

$$\overline{r}^{(k)} = \frac{1}{N_s - 2} \sum_{i=2}^{N_s - 2} r_i^{(k)} \ . \tag{5}$$

The maximum deviation is

$$G_{C_2}^{(k)} = \max_i \left| r_i^{(k)} - \overline{r}^{(k)} \right| . \tag{6}$$

Criterion C_2 is considered satisfied if the obtained values $r_i^{(k)}$ are close to unity.

Criterion (C_3) (convergence rate). The convergence rate criterion reflects the speed at which distances decrease as the signals approach the reference pattern. A relative change indicator is introduced

$$v_i^{(k)} = \frac{d_i^{(k)} - d_{i+1}^{(k)}}{d_{i+1}^{(k)}}, \ i = 1, ..., N_s - 1 . \tag{7}$$

The average convergence rate is defined as

$$\overline{v}^{(k)} = \frac{1}{N_s - 1} \sum_{i=2}^{N_s - 1} v_i^{(k)} \ . \tag{8}$$

The standard deviation is

$$\sigma_{C_3}^{(k)} = \sqrt{\frac{1}{N_s - 1} \sum_{i=1}^{N_s - 1} \left(v_i^{(k)} - \overline{v}^{(k)} \right)^2} \ . \tag{9}$$

Smaller values of this measure indicate a more stable convergence process and, consequently, higher adequacy of the recognition method.

Criterion (C_4) (sensitivity). The sensitivity criterion characterizes the robustness of recognition methods with respect to variations in the displacement step size. Two artificial families of signals constructed with step sizes $h_1$and h_2, where $h_1<h_2$, are considered.

For each criterion C_l, l=1, 2, 3, the following ratio is introduced

$$u_{lk}^{(h_1,h_2)} = \frac{G_{lk}(h_2)}{G_{lk}(h_1)}, \tag{10}$$

where $G_{lk}(h)$ denotes the deviation measure of the l-th criterion computed for step size h. The generalized sensitivity coefficient is defined as

$$u_k = \sum_{l=1}^{3} u_{lk} . \tag{11}$$

A recognition method is considered to satisfy criterion C_4, if the values u_k do not decrease when the displacement step size increases.

These criteria jointly capture complementary aspects of recognition algorithm behavior, including distance regularity, structural symmetry, convergence dynamics, and sensitivity to controlled perturbations in the input data.

The objective of the present study is to empirically verify the robustness and reproducibility of criteria C_1–C_4 using a new artificial family of images constructed from national emblems. To achieve this objective, the following tasks are performed:

1. construction of artificial image series with predefined displacement parameters;
2. extraction of their feature representations;
3. computation of distances D_k;
4. evaluation of indicators (1)–(11);
5. analysis of the compliance of recognition methods with criteria C_1–C_4.

The obtained results are subsequently used to assess the universality of the proposed adequacy criteria under new experimental conditions.

3. Empirical Validation of Adequacy Criteria Using Artificial Image Families

In the present study, a truncated image of the State Emblem of the Republic of Azerbaijan, denoted by I_0, is used as the object of recognition (Fig. 1). The truncation of the original emblem introduces partial structural distortions, which makes the image suitable as a test case for evaluating the robustness of recognition methods under incomplete structural information.

As a reference database, the set

$$B=\{I_1, I_2, \ldots, I_9\},$$

is used. This set consists of full, undistorted images of national emblems from different countries. In particular, the element I_1 corresponds to the complete image of the State Emblem of the Republic of Azerbaijan and is used as the primary reference pattern.

The reference image set B, illustrated in Fig. 2, serves as the basis for constructing artificial image families and for performing subsequent comparative analysis.

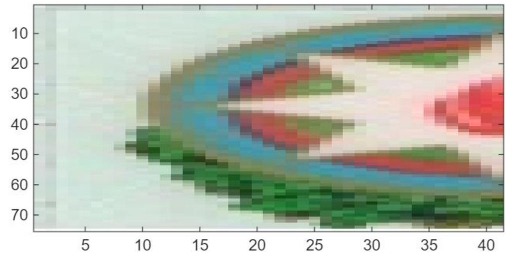

Fig. 1. Truncated image of the State Emblem of the Republic of Azerbaijan I_0.

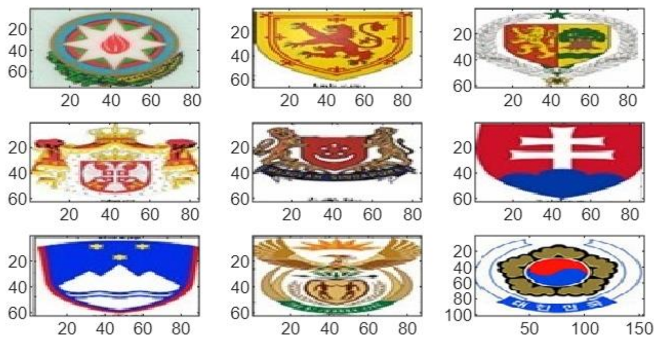

Fig. 2. Reference image set $B=\{I_1, I_2, \ldots, I_9\}$.

The objective of the recognition procedure is to determine the degree of similarity between the image I_0 and each element of the reference database I_j, $j=1, \ldots, 9$, using wavelet-based feature analysis. After preprocessing, wavelet transformation, and feature vector construction for the image I_0 and each reference image I_j, Euclidean distances between the corresponding feature vectors are computed.

Based on these values, a distance matrix

$$D=\{D(I_i, I_j)\} \ (i, j=0, 1, \ldots, 9),$$

is constructed, where $D(I_i, I_j)$ characterizes the degree of similarity between images I_i and I_j represented in the wavelet feature space. The diagonal elements of the matrix are equal to zero and correspond to the distance of an image to itself. The remaining elements reflect the relative similarity between the query image I_0 and the reference images.

The resulting distance matrix is presented in Table 1.

Table 1. Relative similarity measures between image I_0 and reference images I_j ($\times 10^2$)

I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9
0	2.15	2.91	2.79	2.77	2.18	1.64	2.01	2.41	2.49
2.15	0	2.91	2.79	2.77	2.18	1.64	2.01	2.41	2.49
2.15	2.91	0	2.79	2.77	2.18	1.64	2.01	2.41	2.49
2.15	2.91	2.79	0	2.77	2.18	1.64	2.01	2.41	2.49
2.15	2.91	2.79	2.77	0	2.18	1.64	2.01	2.41	2.49
2.15	2.91	2.79	2.77	2.18	0	1.64	2.01	2.41	2.49
2.15	2.91	2.79	2.77	2.18	1.64	0	2.01	2.41	2.49
2.15	2.91	2.79	2.77	2.18	1.64	2.01	0	2.41	2.49
2.15	2.91	2.79	2.77	2.18	1.64	2.01	2.41	0	2.49
2.15	2.91	2.79	2.77	2.18	1.64	2.01	2.41	2.49	0

In practice, it is sufficient to consider only the first row of the matrix, which will be used in the subsequent analysis. The presented distance matrix is symmetric because the Euclidean metric is employed, for which $D(I_i, I_j)=D(I_j, I_i)$. In fact, the matrix is constructed from the base distance vector

$$d = \{D(I_0, I_1), D(I_0, I_2), \ldots, D(I_0, I_9)\},$$

while the remaining elements are obtained by symmetric reflection of this vector with respect to the main diagonal.

According to Table 1, the minimum distance from image I_0 is achieved for image I_6 and equals $D(I_0, I_6)=1.64\times 10^2$. This value is significantly smaller than the remaining elements of the distance vector d.

Thus, within the adopted global similarity measure, image I_6 appears to be the closest match to the query image I_0. However, visual inspection of the corresponding images reveals the absence of pronounced structural correspondence between I_0 and I_6. This discrepancy indicates that the employed global comparison procedure may in certain cases be insufficiently sensitive to semantically significant structural elements of images. Similar limitations of global similarity measures have been discussed in the literature on image quality assessment and structural similarity metrics [12].

To overcome this limitation, an alternative matching approach is introduced that takes into account the spatial organization of structural components. For this purpose, each reference image is partitioned into ten non-overlapping sectors with respect to its geometric center. Such partitioning enables the transition from global comparison to local analysis of spatially aligned regions and allows a more sensitive comparison of structural elements.

As an initial example, the reference image I_1 is selected. Partitioning this image into sectors generates an artificial family of shifted images $F_1=\{F_{1,1}, F_{1,2}, \ldots, F_{1,10}\}$, illustrated in Fig. 3.

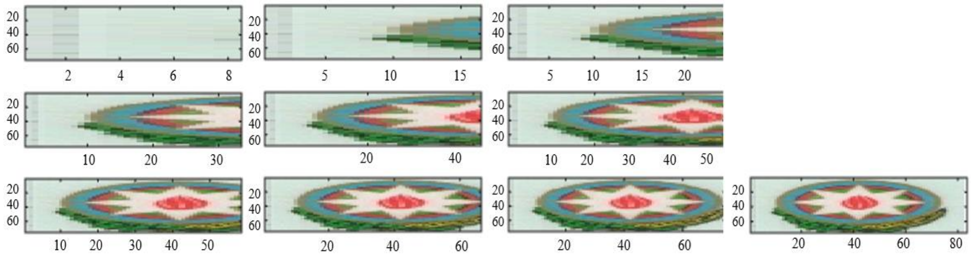

Fig. 3. Artificial image family $F_1=\{F_{1,1}, F_{1,2}, \ldots, F_{1,10}\}$.

The images of the family F_1 are first normalized and segmented. For each segment, a local feature is computed to form a reduced two-dimensional representation. The resulting feature matrices are then linearized according to a fixed rule, transforming each image into a corresponding one-dimensional signal representation.

The obtained family of one-dimensional signals $S_1=\{s_{1,1}, s_{1,2}, \ldots, s_{1,10}\}$ shown in Fig. 4, is used for subsequent wavelet analysis and for evaluating adequacy criteria C_1–C_4.

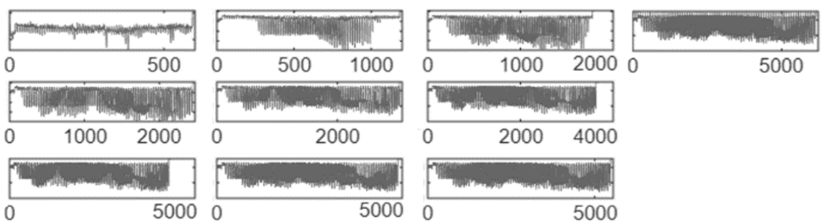

Fig. 4. Family of one-dimensional signals $S_1=\{s_{1,1}, s_{1,2}, \ldots, s_{1,10}\}$ corresponding to F_1

The Euclidean distances between the feature vector of the one-dimensional representation s_0 of image I_0 and the signals from the set S_1 are summarized in Table 2.

Table 2. Similarity measures between image I_0 and images $F_{1,k}$ (k=1–10)

s_0	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}
0.00	336.15	339.33	301.35	177.86	0.00	176.84	224.96	256.47	253.74

From Table 2 it can be observed that the minimal distance corresponds to image $F_{1,5}$, which is consistent with visual inspection.

In a similar manner, using the reference image I_2, another artificial image family $F_2=\{F_{2,1}, F_{2,2}, \ldots, F_{2,10}\}$ is constructed (Fig. 5). After linearization of the images $F_{2,k}$, k=1, 2, …, 10, this family is transformed into the corresponding set of one-dimensional signals $S_2=\{s_{2,1}, s_{2,2}, \ldots, s_{2,10}\}$ shown in Fig. 6.

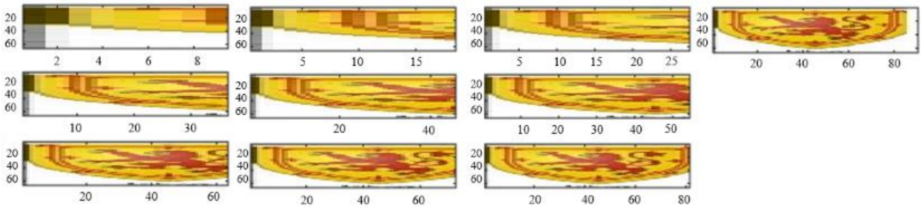

Fig. 5. Artificial image family $F_2=\{F_{2,1}, F_{2,2}, \ldots, F_{2,10}\}$.

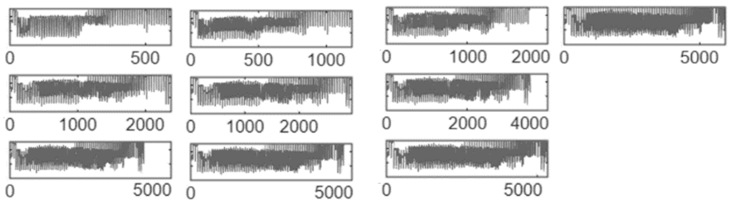

Fig. 6. Family of one-dimensional signals $S_2=\{s_{2,1}, s_{2,2}, \ldots, s_{2,10}\}$, corresponding to F_2.

The Euclidean distances between the signal s_0 and the elements of the set S_2 are presented in Table 3.

Table 3. Similarity measures between image I_0 and images $F_{2,k}$

s_0	s_{21}	s_{22}	s_{23}	s_{24}	s_{25}	s_{26}	s_{27}	s_{28}	s_{29}
0.00	326.86	341.62	303.09	262.92	251.10	266.13	281.29	287.54	295.63

Using the data presented in Tables 2 and 3, the adequacy criteria C_1–C_4 were evaluated according to formulas (1)–(11) for the artificial image families $I_{01}=\{I_0, F_{1,1}, F_{1,2}, \ldots, F_{1,10}\}$ and $I_{02}=\{I_0, F_{2,1}, F_{2,2}, \ldots, F_{2,10}\}$. The obtained results are summarized in Table 4.

Table 4. Results of applying criteria C_1–C_4 to families I_{01} and I_{02}.

Family	Criteria			
	C_1	C_2	C_3	C_4
I_{01}	0.0081	0.5902	0	0.5983
I_{02}	0.0249	0.8674	0.8674	1.7598

The comparative analysis shows that the final score for family I_{01} according to criterion C_4 (0.5983) is significantly smaller than the corresponding value for family I_{02} (1.7598). This indicates a higher adequacy of the recognition results for family I_{01}.

The same procedure was applied to the remaining artificial families $I_{03}=\{I_0, F_{3,1}, F_{3,2}, \ldots, F_{3,10}\}$, $I_{04}=\{I_0, F_{4,1}, F_{4,2}, \ldots, F_{4,10}\}$, ….., $I_{09}=\{I_0, F_{9,1}, F_{9,2}, \ldots, F_{9,10}$. The aggregated results obtained using criteria C_1–C_4 for all families are summarized in Table 5.

Table 5. Results of applying criteria C_1–C_4 to families I_{01} и I_{09}

Family	Criteria			
	C_1	C_2	C_3	C_4
I_{01}	0.0081	0.5902	0	0.5983
I_{02}	0.0249	0.8674	0.8674	1.7598
I_{03}	0.0149	0.7674	0.7674	1.5498
I_{04}	0.022	0.8464	0.8464	1.7148
I_{05}	0.011	0.6771	0.6771	1.3651
I_{06}	0.0079	0.6351	0.6351	1.278
I_{07}	0.0168	0.8132	0.8132	1.6433
I_{08}	0.009	0.5975	0.5975	1.2041
I_{09}	0.0111	0.7371	0.7371	1.4852

As can be seen from Table 5, the smallest aggregated score is obtained for family I_{01}. The fact that the recognition object belongs to this family further confirms the adequacy and reliability of the proposed evaluation criteria.

4. Conclusion

This paper presents an extended empirical verification of adequacy criteria for evaluating image recognition methods using artificially generated families of national emblem images. The study was aimed at analyzing the reproducibility, robustness, and universality of criteria C_1–C_4 under conditions of a strictly controlled computational experiment. Artificial image families and their corresponding one-dimensional signal representations were constructed using a sector-based partitioning strategy combined with signal linearization and feature extraction based on the continuous wavelet transform. Such an approach enables systematic analysis of algorithm behavior under controlled spatial displacements and structural variations.

The experimental results obtained for several artificial image families demonstrate that the proposed criteria consistently identify the most adequate recognition outcomes. In particular, the lowest aggregated scores are observed for families corresponding to visually and structurally consistent patterns, which confirms the reliability of the proposed evaluation framework. The conducted analysis also reveals limitations of conventional global similarity measures, which in some cases fail to capture semantically significant structural correspondences between images. The proposed sector-based matching scheme mitigates this limitation by incorporating local spatial information into the recognition process.

Overall, the obtained results confirm the applicability of the adequacy criteria to a new class of structurally complex images and demonstrate their robustness with respect to variations in feature extraction and image representation procedures. The developed experimental framework provides a reproducible and interpretable tool for systematic evaluation of image recognition methods.

Future research will focus on extending the proposed approach to larger and more diverse image datasets, exploring alternative feature representations, and integrating statistical significance analysis in order to further strengthen the methodological foundation of adequacy evaluation.

References

1. Bay H., Tuytelaars T., Van Gool L. SURF: Speeded up robust features, Comput. Vision Image Underst., V.110, N.3, (2008), pp.346–359.
2. Christopher M. Bishop, C.M. Pattern Recognition and Machine Learning. New York: Springer, (2006).
3. Daubechies I. Ten Lectures on Wavelets, Philadelphia: SIAM, (1992).
4. Duda R.O., Hart P.E., Stork D.G. Pattern Classification. 2nd edn. New York: Wiley, (2001).
5. Islam S.M., Debnath R. A comparative evaluation of feature extraction and similarity measurement methods for content-based image retrieval, Int. J. Image Graphics Signal Process, V.12, N.6, (2020), pp.19–32.
6. Lowe D.G. Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision, V.60, N.2, (2004), pp.91–110.
7. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd edn. Burlington: Academic Press, (2009).
8. Mohammadi P., Ebrahimi-Moghadam A., Shirani S. Subjective and objective quality as-sessment of image: A survey. arXiv:1406.7799 (2014).
9. Rehman A., Gao Y., Wang Z., Wang Q. Image classification based on complex wavelet structural similarity,Pattern Recogn., V.46, (2013), pp.542–554.
10. Rzayev R.R., Kerimov A.B., Garibli U.G., Salmanov F.M. Criteria for Assessing the Adequacy of Image Recognition Methods and Their Verification Using Examples of Artificial Series of Signals, Problems of Information Society, V.15, N.1, (2024), pp.10–17.
11. Rzayev R.R., Kerimov A.B., Aliyev V.R. Application of Machine Vision Technology for Multicopter Flight Control Under Active Operation of Counter-UAV Systems. In: Intelligent Computing, Proceedings of the 2025 Computing Conference, V.3, (2025), pp.88–102.
12. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process, V.13, N.4, (2004), pp.600–612.