IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 February 2025, accepted 4 March 2025, date of publication 6 March 2025, date of current version 14 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3548910

RESEARCH ARTICLE

Image Description Compression in Classification Structural Methods

VOLODYMYR GOROKHOVATSKYI[1], IRYNA TVOROSHENKO[1], OLENA YAKOVLEVA[1,2,3], AND MONIKA HUDÁKOVÁ[2]

[1]Department of Informatics, Kharkiv National University of Radio Electronics, 61166 Kharkiv, Ukraine
[2]Department of Economics and Finance, Economics and Management Institute, Bratislava University of Economics and Management, 851 04 Bratislava, Slovakia
[3]Department of AI Research and Development, SYTOSS Ltd., 831 04 Bratislava, Slovakia

Corresponding author: Iryna Tvoroshenko (iryna.tvoroshenko@nure.ua)

This work was supported by the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under Project 09I03-03-V01-00115.

ABSTRACT The problem solved in the article is reduction of computational costs for the image classification process when applying structural methods. The main focus is implementing tools for granulation, screening, and clustering processing a set of elements of etalon descriptions. As a result of compression, each etalon is transformed into a reduced set of descriptors or data centroids, ensuring high speed and performance of image classification. Several variants of simple data compression schemes are assessed and compared to the traditional linear search method, along with two variants of etalon clustering. The comparison includes results achieved for the entire data set and for each of the images separately. The paper presents the results of software modeling of the proposed approaches for two experimental sets containing images of football club logos and artistic paintings. The test sample includes a set of images from the etalon database along with other images that do not belong to the database and with a set of geometric transformations of shift, scale, and rotation in the field of view applied to them. The research covers practical issues of choosing threshold parameters to set the equivalence of descriptors and minimizing the number of class votes to ensure the required level of classification accuracy. Testing has confirmed a significant processing acceleration and a sufficiently increasing level of classification accuracy due to employing compression. Particularly the conducted modeling revealed a tenfold increase in speed. It has been experimentally confirmed that using a clustering apparatus has a much higher potential in terms of classification accuracy and speed than simple sifting or granulation schemes based on close description components.

INDEX TERMS Classification speed, data clustering, description compression, granularity, image classification, keypoint descriptors.

I. INTRODUCTION. LITERATURE REVIEW

One of the most significant, relevant and still unsolved problems in computer vision systems is ensuring the required performance of their applied functioning, particularly, the accuracy and speed of data processing. This problem is especially applicable to visual object classification tasks where a large number of prototype classes are used, the images of which are represented as a set of vectors. Traditionally, such tasks are referred to Big Data [1], as they require a volumetric linear search in a multidimensional data space [2], [3], [4], [5], [6].

Among applied applications, the most widespread is the continuous-group approach to recognition based on a deterministic model of forming the variety of analyzed images [6], [7]. According to this model, an endlessly powerful class of images is generated using a set of priori unknown topological transformations applied to some fixed etalon realization [7], [8], [9].

Modern structural methods classifications analyze a visual object as a set of keypoint descriptors [7], [9], [10], [11], [12]. Special filters – detectors, calculate keypoint descriptors. The

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

keypoint descriptor is represented as a binary vector of 256 or 512 elements [6], [11], [12].

Structural methods use the bag-of-words model to process classification, and the result of voting for the set of descriptors of the recognized object determines the class of an object. Consider that structural analysis has certain advantages over modern neural network technologies, as it is based on direct alignment with the description or parameters of the etalons and does not require long-term training [3], [5], [6]. Moreover, applying structural methods allows quick changes in the etalon data set content and composition, making these methods universal regarding the set of recognized images [13]. Another significant advantage of the implementation of structural methods is the ability to make a decision and classify parts or details of the description of a visual object. The functionality of the keypoint descriptors apparatus has the property of invariance to geometric transformations of objects in the field of view, which is essential for solving applied tasks [5], [6], [7], [8], [9].

The use of structural methods for image classification claimed as more efficient than artificial neural networks, as they allow to avoid such problems as the difficulty of constructing effective network training algorithms [14], significant training time [15], determining the productive network dimension, high requirements for computing resources [16], the need to ensure invariance to geometric transformations of objects, the limited range of tasks for a particular network solely based on the training result, etc. [10], [17], [18], [19].

Recently, deep learning methods [20] based on statistical characteristics of the image texture have shown decent efficiency for the task of image classification using neural network methods. However, the effectiveness of these methods, like all neural network approaches, often depends on the synthesized network structure, which, in turn, is related to the application and observation conditions. At the same time, the task of forming a compressed set of features does not lose its relevance when implementing a deep learning apparatus.

At the same time, structural methods provide possibility for improvement by reducing the computational cost of classification [3], [20], [21], [22], [23]. One of the improvement options is to develop hierarchical methods of data retrieval based on clustering, hashing, or evaluation [8], [20], [21]. There are also ways to reduce the dimensionality of the structural description by filtering out descriptors based on the criterion of information content [8] or uniqueness [20]. For example, the informativeness criterion is based on the value of the difference in the distance of the descriptor to its own and other etalons in the classification database. However, the key factor for effective classification is still the group metric correlation between sets of descriptors [3].

Enhancing the performance of structural methods can also be achieved by reducing the number of descriptors in the description of the etalons [24]. Any reduction in the number of descriptors is crucial because it helps skip difficultly implementable combinatorial aspects while solving the task of comparing multidimensional feature sets [24], [25].

Based on the mentioned above advantages, the application of the principles of data granulation can be considered a promising solution [4], [22], [26], [27], [28]. The founding of granulation in computational intelligence with the introduction of fuzzy set apparatus belongs to Zadeh [27], [29]. Granulation is the process of obtaining aggregated data that is significant in some sense [4], [26].

In particular, granularity can be performed according to the indistinguishability criterion, which states that components should be grouped together if they demonstrate similarity. Describing data in different degrees of detailization can be used to contribute to the creation of hierarchical systems with multiple levels of granularity [3], [4], [22]. However, achieving the desired level of abstraction based on the task and data description requires experimental validation.

One of the methods of description reduction that have been studied is determining the set of centers constructed on the basis of percentile relations concerning the medoid for the data set [21]. Other developed approaches implement the removal of a subset of descriptors with low informative value [3], [8].

An effective example of implementing the granulation principle for integrating in the space of image features is the pyramidal representation [6], [7], [9]. The pyramidal structure integrates the brightness for spatially close image pixel coordinates and reduces the amount of multi-scale visual information. The concept of data granularity has become widespread in pattern recognition systems. Various aspects of creating an effective feature space in image classifiers are studied [2], [5], [6], [30]. Granularity is used in the object-background segmentation methods to improve performance [8], [10]. Effective metrics in the granular data space have been presented [31]. The outcomes of applied research on the use of compressed description representation have been obtained [8], [21], [32], [33], [34].

Methods of clustering, hash baskets, or center estimation can be considered metric granulation, as they are based on calculating the proximity of data to each other using the metric [3], [4], [8], [28], [35]. Implementing these approaches improves processing speed with an acceptable decrease in classification accuracy. However, achieving the required performance values requires a deeper study of all the possibilities and nuances of the applied use of the granulation and sieving apparatus to fulfill the need for objective confirmation of their universality in the field of analysis of multidimensional data.

The range of metric granulation methods is quite extensive, from relatively simple approaches related to identifying pairs of the most similar description descriptors to discarding one of these pairs. At the same time, this range includes the most

complex multistage clustering methods in multidimensional data spaces [3], [35].

For example, similarity (equivalence) granulation involves finding the metrically closest descriptor and excluding it from further processing. This step helps to reduce the amount of computation at times while maintaining the representation of descriptors in the description.

As an alternative to intelligent methods, reducing the description through simple schemes of eliminating elements according to some law of order in the full description can be also considered. Sifting involves the exclusion of, e.g., every second element from the dataset, which makes it possible to halve the data volume and reduce the amount of computation. It is easier to perform, does not require distance estimation, and is fast. Another scheme is to generate a small description of the keypoints directly by the detector or to randomly select a reduced subset from the full description.

Definitely, these simplified transformation methods do not consider the significance and spatial relationships of the elements, which can lead to the loss of important information and a decrease in classification accuracy.

This research and paper focuses on the study and experimental implementation of metric granulation and data sifting in structural methods classification, as well as on the analysis of their effectiveness compared to the traditional method without granulation. The main goal is to study the impact of the proposed modifications on the speed and accuracy of classification, which is important for improving the performance of computer vision systems in practical applications.

The article's contribution is a detailed study of the effectiveness of creating a hierarchical description representation with several hierarchy levels. Another outcome is the confirmation of the practical advantages of implementing intelligent analysis by creating granular descriptions compared to trivial sifting schemes and generating abbreviated descriptions directly by keypoint detectors.

II. FORMULATION OF METRIC GRANULATION

Let us present a structural description of $Z = \{z_i\}_{i=1}^{s}$ visual object (a set of keypoint descriptors) as a set of $F(Z)$ transformed descriptions f_r, obtained at successive levels of hierarchical representation

$$F(Z) = \cup \{f_r\}_{r=1}^{R}, s(r) = card\, f_r, \tag{1}$$

where $s(r)$ is the power of the description of f_r at the hierarchy level with the number r. At $r = 0$, there is the lowest level, representing the full description of Z. The highest level with the number R is the transformation in the form of a single data vector. The medoid of a set of vectors or center-vectors including averages of set's elements may be an example [8], [20]. The hierarchies of f_r are formed sequentially in a recursive way, starting from the lowest level (description Z). The result is a structured sequence of descriptions

$$Z \to f_1 \to f_2 \to \ldots \to f_R.$$

The purpose of constructing a hierarchy is to reduce the dimensionality of the description and improve classification performance by speeding up data processing procedures. To create a consistent hierarchy of descriptions, at each step, a specific procedure T of transformation is applied to reduce the power of the description when moving to the next level

$$f_{r+1} = T\{f_r\}, s(r+1) \ll s(r).$$

The task of forming a higher-level hierarchy is set until the classifier using the compressed description of f_{r+1} will effectively produce the result of a given accuracy. And the main task of transforming T is to speed up the classification process by generalizing the description data.

The definition of T is a data granulation procedure based on the model of forming one vector of the next level based on the analysis of a pair of the most similar vectors of the previous level. This action is used in the process of data clustering [8], [28], [35].

Recursion allows the application of the same procedure that establishes the law of transition to the next level. This procedure delivers one element of the upper level based on a group of elements of the lower level. This is a model of pyramidal recursive structures. In fact, the procedure T splits the set of descriptors into pairs of equivalent elements.

With the introduction of T to describe the etalon E_k with the number of s descriptors, the following transformation is obtained

$$T(E_k) \to E_k^*, card\left(E_k^*\right) = s(k) \ll s, \tag{2}$$

which results in a filtered description of E_k^*, that is a subset of $E_k^* \subset E_k$ elements from E_k, selected according to a certain granularity criterion. For better practical results, it is proposed to reduce the description power by about 1.5–2 times at each processing stage. As a result of the implementation of (2), the $E = \cup E_k$ data set acquires a shortened representation as $E^* = \cup E_k^*$.

The criterion in (2) is the definition of the next level element through a pair of the closest elements of the lower level. Thus, the metric proximity plot is the basis for forming the next level granule.

The principle of analyzing the metric proximity of descriptors can serve as the basis for creating various options for filtering (reducing) data. Considering specific options for constructing the T transformation, based on the value of the ρ metric for descriptors, any logical operation can be obviously used to obtain the next-level granules.

A. GRANULATION BY EQUIVALENCE THRESHOLD

For each of the descriptors $e_d(k) \in E_k$ the closest descriptor $e_*(k)$ in the set E_k is determined by the metric ρ. And it corresponds to the linear metric search model

$$e_*(k) = \arg\min_{v \neq d} \rho(e_v(k), e_d(k)). \tag{3}$$

After determining the descriptor with the minimum metric, the equivalence condition is checked: $\rho(e_d(k), e_*(k)) \leq \delta_\rho$,

where δ_ρ is a given threshold parameter for the metric value, which approximately establishes the equivalence of two descriptors. In the space of multidimensional vectors, the value of δ_ρ is often set to 25% of the maximum value of the metric, but it can be changed depending on the task and input data [8], [22].

If the equivalence by the δ_ρ threshold in (3) is not established, then the first argument is recorded in E_k^*, this element is excluded from further processing, and the second one continues to participate in the analysis. It's worth mentioning that the order of descriptors in the description and the value of δ_ρ directly affect the results of granulation.

In fact, the model (3) for determining the equivalence of two vectors is supplemented by the predicate P:

$$P(z_1, z_2) = 1 \mid \rho(z_1, z_2) \le \delta_\rho. \tag{4}$$

The number of vectors obtained using this variant depends on the threshold δ_ρ.

An alternative way to implement (3) is to determine the pair of nearest vectors without applying the condition (4). In this case, the amount of description is reduced level by level by about half. If the closest one with a threshold is determined, the advantage is less than 2 times, and the specified distance limit between the newly created granules is preserved, which maintains the degree of differences for the generated data. There is even simpler analysis option to search with a specified equivalence threshold until the first identical element is determined.

B. GRANULARITY WITH THE SMALLEST DISTANCE FILTERING

To calculate the distance matrix between all description descriptors E_k [5], the selection method starting with the smallest distance is applied. The next descriptor is added to the list E_k^*, and the corresponding row and column of this descriptor are excluded from the matrix. The next step is analyzing the descriptor with the next largest distance value. This way a fixed number of descriptors is consistently added to the set E_k^*. This method is a bit more complicated than the previous one, as it may include sorting by distance. However, it does not depend on the order of descriptors and does not require adapting the δ_ρ threshold to the content of the data.

As a result, the lower-level features of the researched analysis methods are combined into complexes and further considered as a single feature. Compared to clustering, the studied granulation procedure creates a modified data space controlled by the equivalence parameter step by step.

As an alternative to the granularity methods discussed above, reducing the description by eliminating elements with a fixed step is considered. For example, the method of selective exclusion may be applied, where every second descriptor is removed from the set regardless of its characteristics.

The proposed method has a number of well-known drawbacks. The loss of important descriptors that are critical for classification may be one of flaws. Discarding elements is done mechanically without assessing their importance or role in the overall structure of the set. After such rejecting, either only important or only unimportant descriptors may remain. For complex tasks requiring high accuracy, this method cannot be effective, as it does not provide in-depth data analysis and may lead to incorrect results.

III. CLUSTERING FOR A SET OF DESCRIPTION DESCRIPTORS

Clustering is referred to as a data quantization apparatus with the definition of a system of centers [8], [28], [35]. The usage of the clustering apparatus on a set of descriptors is studied.

The database E of etalon descriptions will be clustered. Two alternative approaches to clustering are researched: one with the same number of M clusters – collectively for the entire set E (system of centers W^1) and separately for each of the etalons E_i, $i = 1, \ldots, N$ (system of centers W^2, m centers for each etalon) [8]. As a result, two options for centers are obtained

$$W^1 = \left\{w_j^1\right\}_{j=1}^{M}, W^2 = \left\{w_j^2\right\}_{j=1}^{M}. \tag{5}$$

The equal number of M clusters in (5) is chosen solely to ensure an objective comparison of the performance of the separate and joint cluster representation of the etalon descriptions. In fact, the set of centers in (5) is an approximation for the set E and creates a new space of classification features.

At the same time, to obtain a vector representation for each etalon according to the constructed system of centers in the first variant (joint clustering for E), it is necessary to perform a competitive analysis of the description for each of the etalons E_k. To do this, for the elements of $z \in E_k$ the number of v of the nearest center in the set W^1, $w_j \in W^1$ is determined

$$v = \text{argmin}_{j=1,\ldots,M} \rho\left(w_j, z\right). \tag{6}$$

The scheme (6) is an embodiment of the search model for the nearest cluster center. As a result of applying (6), a representation of the etalon E_k in the form of an integer vector $gE_k = [g_1, g_2, \ldots, g_M]$ is obtained with the number of components equal to the number of clusters. Here, g_d is the number of description descriptors of E_k that were assigned to the cluster with the number d according to the model (6). The vector g is the distribution of the description data over the set of clusters. In the process of classification, for the analyzed image with the description Z, it is likewise necessary to obtain the vector $g(Z)$ of the cluster representation.

For the system W^2 a fixed equivalent number m of cluster centers for each of the etalons is already obtained, and $M = mN$, since the etalons were clustered separately. Therefore, no additional analysis or transformation of the data etalon representation is required. In general, the number of centers m is a parameter that depends on the data composition.

IV. CLASSIFICATION SCHEMES AND CRITERIA

The technology R of structural classification is implemented in two stages according to the model $R = R_2R_1$. The stage

R_1 executes a local decision regarding the class of object descriptors, and the stage R_2 based on the decisions (votes) of the components of the entire description, receives the class value for the analyzed object [5], [8].

The classifier R_1 applies an element-by-element analysis to the input description $Z = \{z_v\}_{v=1}^{s}$ of the object and assigns the descriptor $z_v \in Z$ to one of the classes according to the rule

$$R_1 : z_v \to \{1, \ldots, N\}. \tag{7}$$

The class k of the analyzed descriptor z_v of the object in the model (7) is defined as the argument of the minimum distance ρ on the set E of descriptors of all classes

$$R_1 : k = \text{argmin}_{i=1,\ldots,N;\ d=1,\ldots,s}\ \rho\left(z_v, e_d\left(i\right)\right). \tag{8}$$

Here $\rho : B^n \times B^n \to [0, \infty]$ is the distance in the vector space B^n.

In the competitive model (8), it is advisable to use the Hamming distance, which determines the number of distinct bits in two binary vectors [5], [8]. The expression (8) minimizes the distance value by the variable i (class number) on the set of descriptors of the dataset E. In the formula (8), the determination of the minimum can be organized both for the full list $s \cdot N$ of base descriptors and separately for each of the N etalons (with a set limit on the minimum value).

In the separate analysis method, the number of object descriptors corresponding to a fixed etalon is counted for each class. This model applies multiple vote counting when the same object descriptor can find a match for the etalons of different classes.

Thus, model (8) implements multiclass classification for a single descriptor in the vector space B^n based on the best match on the set E of etalons.

To implement R_2, a vector $\{h_i\}_{i=1}^{N}$ is introduced with integer values for the accumulated number of class votes obtained by applying (8) to the object Z. Based on R_1, for each $z_v \in Z$, according to (8), the class number k is determined, and then the vote accumulator $h_k = h_k + 1$ for the corresponding class number is incremented. The resulting vector $\{h_i\}_{i=1}^{N}$ is the distribution (histogram) of the votes number of the set Z in the class system.

As a result of processing the object description Z, the vector $\{h_i\}_{i=1}^{N}$ is calculated. The class Z is defined by the rule R_2 as the maximum argument

$$R_2 : Z \to E_k | \left(k = \arg\max_{i=1,\ldots,N} h_i \right) \& \left(\mathrm{h_k} \geq \delta_h\right), \tag{9}$$

where δ_h is a threshold for the minimum number of votes. If the condition $(\mathrm{h_k} \geq \delta_h)$ is not met, the object's class is considered unassigned. This is a refusal to classify when there is no reason to assign an object to any of the classes.

The value of δ_h is determined experimentally for the database E. As a rule, δ_h is chosen as the minimum number of votes (with a tolerance) required to confidently classify the test set based on the etalons. From the general theoretical point of view, δ_h is a compromise in the task of distinguishing the etalons from the rest of the images that can be input to the classification system [3], [8].

The sequence of rules R_1, R_2 implements a classifier based on solutions obtained for a set of object components. It implements the principle of structural analysis and provides resistance to spatial distortions of descriptors due to possible interference.

The traditional classification method implements models (8), (9) for the full set of database descriptors. The classification process using granular descriptions is modified by performing a limited search procedure on the reduced database $E^* = \cup E_k^*$. It is necessary to consider that the description of the object remains complete, while the descriptions of the etalons are reduced [8]. It is supposed that the full-fledged E_k and transformed E_k^* etalons have approximately equivalent power values for different classes. In this variant s elements of an object distribute their votes according to the system of N classes, and as a result, the following is obtained $\sum_{k=1}^{N} h_k \leq s$.

The value of δ_h is determined experimentally for a given database E of etalons. As a rule, it is determined as the minimum number of votes h_m (with a tolerance), required for confident classification of the test sample based on the etalons. From a general theoretical point of view, δ_h is a compromise in the task of separating the etalons descriptors from the descriptors of the vast majority of other images that may be input to the classification system [8].

In practical application, choosing a threshold based on the $\delta_h = 0.95 h_m$ model is recommended, which corresponds to a 5% downward deviation from the h_m minimum.

The classifier using joint clustering based on the system of centers W^1 first receives the vector representation $g\left(Z\right) = [g_1, g_2, \ldots, g_M]$ of the recognized object according to the model (6), and then, using the metric μ for vectors of integers, determines the class of the object as an argument to the minimum of the metric μ on the set of cluster representations for the base etalons E

$$R_2 : Z \to E_v | v = \arg\min_{k=1,\ldots,N} \mu\left(g\left(Z\right), g\left(E_k\right)\right). \tag{10}$$

A variant of the metric μ in (10) can be the Manhattan distance of the space of integer vectors. Using the rule (10) instead of (8), (9) for the traditional method gives a boost in computational volume approximately proportional to the value of $\gamma = s/M$. Note that in order to filter out images outside the database, the optimal value of the metric μ in (10) must satisfy the significance condition $\mu_v \geq \delta_\mu$.

The classifier based on separate clustering at the stage R_1 implements the model (6) on the set of centers $w_j^2 \in W^2$. In this case, the class of the descriptor is uniquely determined by the identified cluster center, and the class of the analyzed object is estimated by the model (9). The enhancement of computational time compared to the traditional method is also proportional to the value of γ.

When local interference is present, the classification using the methods under consideration is actually based on the

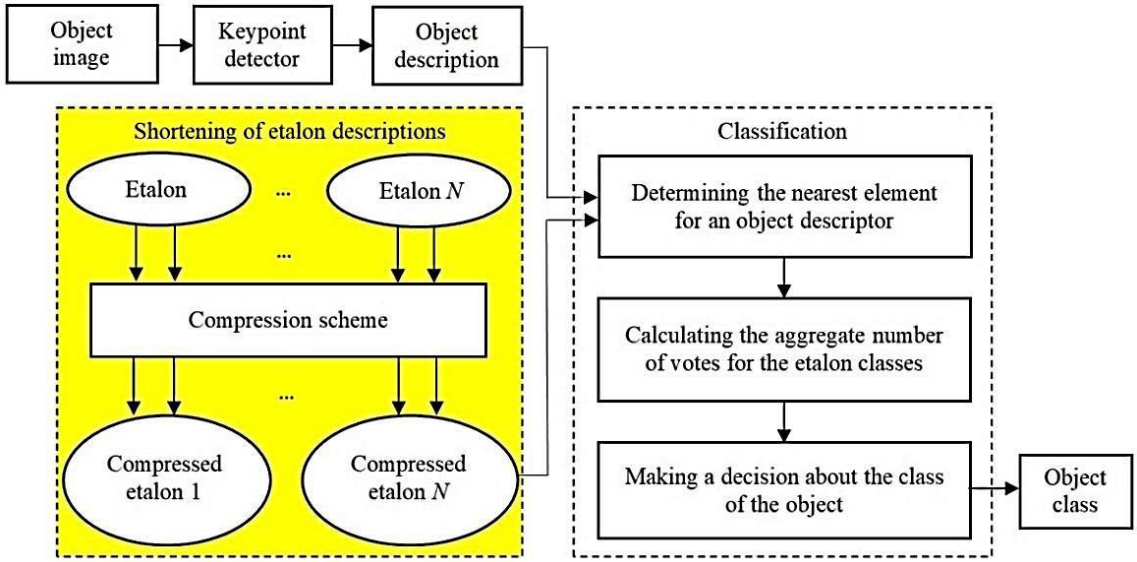


FIGURE 1. Classification scheme based on the abbreviated description of the etalons.

undistorted part of the description of the analyzed object. In this case, the number of votes naturally decreases, so the parameters of δ_ρ, δ_h should be adapted to the acceptable level of interference.

Since in separate clustering the approximation is performed by forming centers for the etalon data, the criterion that can predict the classification performance is the average distance between the systems of centers for pairs of etalons. If the vectors for the centers of a pair of etalons is denoted as $A = \{a_i\}_{i=1}^m$ and $B = \{b_j\}_{j=1}^m$, then this criterion β has the form

$$\beta(A, B) = \left[\sum_{i=1}^m \sum_{j=1}^m \rho(a_i, b_j)\right]/m^2. \quad (11)$$

The value of (11) is within the definition of the metric ρ. For binary vectors of dimension n and the Hamming metric, the indicator (11) belongs to $[0, n]$. The larger the value of β is, the greater the dispersion between the systems of formed centers is and the higher the degree of distinction between the classes of A, B in the space of cluster centers is.

The classification scheme, according to the abbreviated description, is shown in Figure 1.

V. EXPERIMENTAL PLAN AND PERFORMANCE CRITERIA

In the experimental part of the paper, the performance indicators (accuracy, speed) are studied for the following typical description reduction schemes:

1) Classification by full descriptions without abbreviations (traditional scheme).
2) Implementation of granulation at the 5th level with a threshold.
3) Implementation of granulation at the 5th level without using a threshold.
4) Formation of a fixed number of 15 descriptors for each etalon image by the Oriented FAST and Rotated BRIEF (ORB) detector [36].
5) Sifting by selecting every 33rd descriptor from the etalon description.
6) Sifting five times by discarding one at a time until reaching 15 descriptors in each description.
7) Selection of the first 15 out of 500 descriptors for each etalon description.

The parameters in schemes 2–7 are chosen to obtain an approximately equivalent number of 15 descriptors in the abbreviated description.

The classification performance is evaluated by the accuracy criterion pr, which reflects the ratio of the number of u test images with a correct class assessment (or definition as an image outside the etalon database) to the total number of U experiments [3], [5]

$$pr = u/U. \quad (12)$$

Another important criterion is the credibility of the classification decision based on the number of votes received by the classes. The credibility of β characterizes the degree of confidence in the decision and is assessed by the criterion in the form of the ratio of the local (second largest) maximum h_{m2} among the votes $\{h_i\}_{i=1}^N$ to the main maximum h_{m1}:

$$\beta = h_{m2}/h_{m1}. \quad (13)$$

The lower the value of (13) is, the higher the credibility of the classification solution is.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The main goal of conducting the described experiments is to confirm the performance and effectiveness of the proposed granulation methods (including clustering) in the



FIGURE 2. Images of coats of arms from the etalon database.

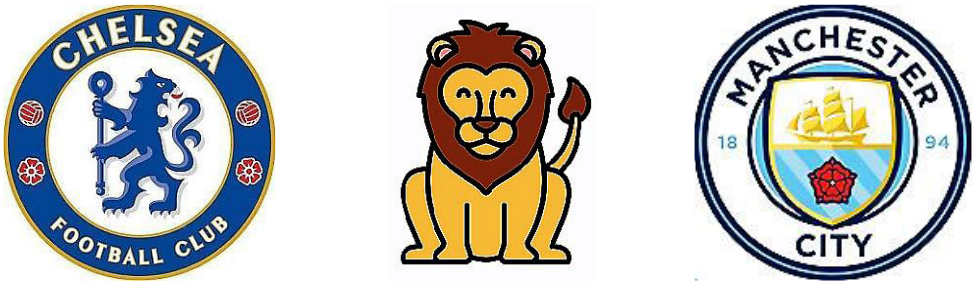


FIGURE 3. Images outside the base.

classification task in comparison with the traditional (full description) and simplified (after discarding) methods.

A. EXPERIMENTS WITH THE DESCRIPTION GRANULATION

The OpenCV library [37] was used to conduct the research. The ORB detector [11] was used to detect and describe keypoints in the images, and sets of keypoint descriptors were determined for the test set of images. The full descriptions of the etalon and test images include 500 ORB descriptors with a possible slight deviation.

The test set in the experiment consisted of 72 images. It contained five etalon images of football club logos (classification base, class alphabet, Figure 2), three out-of-base images (Figure 3), and images obtained from these eight images by applying geometric transformations of scale (10% reduction or increase), rotation (30 degrees to the left or right), and a combination of scale and rotation (Figure 4). Evidently, the first and third out-of-base images selected (Figure 3) have significant visual similarity to some of the database etalons. Figure 4 shows examples of test images with coordinates of keypoints and under the influence of geometric transformations.

A classification error was recorded when a transformed etalon image was assigned to an "alien" class and a transformed image outside the database was assigned to one of the available classes.

To perform the equivalence threshold granulation, the threshold $\delta_\rho = 37.5\%$ was set from the maximum of the metric (256) at each level. The granulation was conducted in several stages, and according to the modeling results, it was decided to stop at the fifth level of the hierarchy, where the level of accuracy begins to decrease significantly. Using the threshold analysis, the number of elements was reduced by less than half at each stage.

TABLE 1. Classification parameters using granularity.

Granulation level	Time, s	Number of base descriptors	Number of votes per class
Full description	29.45	**2492**	500, 500, 495, 497, 500
I	15.19	**1279**	411, 409, 435, 451, 441
II	10.05	**830**	304, 331, 404, 417, 368
III	7.27	**586**	242, 267, 390, 373, 306
IV	5.55	**449**	212, 237, 368, 318, 256
V	4.48	**376**	213, 217, 343, 267, 227

Table 1 shows the data demonstrating the number of database descriptors remaining after performing granularity at each level. Table 1 also contains the number of votes for the correct class and the estimated time for classifying one etalon image at different granularity levels. The results of Table 1 are obtained for the descriptions of the existing etalon database (Figure 1).

As seen from Table 1, as the level of granularity increases from I to V compared to the full description, the number of descriptors in the abbreviated descriptions of the etalons decreases from 2492 to 376, and the classification time decreases proportionally. Even with a significant reduction in the number of informative elements at the Vth granularity level (six and a half times compared to the full description!), the classification accuracy obtained in the experiment



FIGURE 4. Examples of test images with keypoint coordinates.

TABLE 2. Accuracy and credibility indicators for different reduction schemes.

Scheme	1	2	3	4	5	6	7
Accuracy, pr	1.0	0.75	0.63	0.63	0.95	0.85	0.5
Credibility, β	0.20	0.48	0.60	0.57	0.63	0.62	0.72

remains at the highest level $pr = 1$, and the maximum votes in all experiments correspond to the "correct" class.

To illustrate with the concrete example, the values of the maximum votes $\{h_i\}_{i=1}^{N}$ for one of the options for shortening descriptions were 175, 148, 164, 172, 192, which means that according to this data, the threshold δ_h is equal to $\delta_h = 0.95 \times 148 = 141$ votes. So, the further decision to assign a class is made if the number of votes is 141 or more. For images outside the database, the number of votes must be less than 141.

In contrast to the equivalence threshold granulation, the fixed step discarding approach always reduces the number of elements by a corresponding number of times. For example, the number of descriptors when sifting through one at the hierarchy levels was discretely reduced as 2492, 1245, 622, 309, 154, 75.

An important point is to study the impact of geometric transformations on classification performance. The simulation on the test set of 72 images with the implementation of geometric transformations brought the following outcomes. For the experimental evaluation, the results of granulation No. 2 and No. 3 at the 5th level were selected, where the number of descriptors in the description of each etalon was reduced to 15, i.e., 33 times less compared to the full description. The alternatives with the same parameters were the screening schemes No. 4–No. 7 and the traditional method No. 1 with full descriptions of the etalons.

The obtained values of accuracy and credibility are shown in Table 2.

The analysis of the data in Table 2, demonstrates that the traditional method (Scheme 1) based on the full description has the best accuracy (highest value) and credibility (lowest). However, the other methods (Schemes 2–7), which are aimed at speeding up the classification, require approximately 33 times! less processing time and have slightly worse performance.

Granulation with a threshold (Scheme 2) showed slightly better results for both criteria than granulation without a threshold (Scheme 3). This difference can be explained by the introduction of an additional analysis for the value of the minimum distance.

Scheme 7, which uses the first 15 of the generated 500 descriptors from the etalon descriptions, has the worst accuracy and credibility. Schemes with sifting 5 and 6 revealed a fairly high level of accuracy but with a low credibility coefficient. This fact emphasizes that the effectiveness of the discarding procedures (Schemes 4–6), in general, depends significantly on the sequence of descriptors in the description and on the discarding principle itself. To overcome this problem, may require additional multiple processing to achieve the required performance indicators.

Scheme 4 reduction is based entirely on the properties of the ORB descriptor; its implementation does not guarantee high classification performance.

B. EXPERIMENTS WITH DESCRIPTION CLUSTERING

To make a comparison, a software modeling of the classifier was conducted using the two clustering methods considered: for the entire database and separately for each of the etalons. The database containing five images of paintings by the Ukrainian artist M. Primachenko [38] was used, along with three of her other paintings outside the classification database.

Figure 5 shows the paintings from the database used and the coordinates of the keypoints of the descriptors obtained by the Binary Robust Invariant Scalable Keypoints (BRISK) detector (a 512-bit vector) [12]. The artist uses an original technique for her paintings, which at the same time makes them visually similar to each other.

For clustering, $M = 15$ clusters for the database and $m = 3$ for each of the etalons were applied. The conditions of modeling under the influence of geometric transformations corresponded to the methods are described above. The following research results are worthy of mention. The traditional method for images of paintings also showed the highest accuracy $pr = 1$, successfully classifying all images from the database and separating out-of-base images.



FIGURE 5. Images from the image database and coordinates of keypoints.

The modified method based on cluster centers for the full database demonstrated an accuracy of $pr = 0.82$, and with separate clustering – $pr = 0.90$. Notable that the separate clustering gives a slightly higher level of accuracy.

When the number of clusters was increased to $m = 5$ for each etalon (25 centers in total for the database), the accuracy of the method with clustering of the full database naturally increased to the level of $pr = 0.88$, and with separate clustering – to $pr = 0.98$.

Consider that the value of the parameter $m = 5$ in the clustering methods means that only 5 centers of data are used for each of the etalons. In contrast, the data in Table 2 (Schemes 2–7 for the reduction) are obtained for the case when 15 descriptors represent each etalon. Thus, the clustering methods achieve significantly better accuracy (0.88 and 0.98) and three times less computational cost.

Measurements of the time to classify one image in the clustering experiment reveal advancement over the traditional method of about 158 times for the variant $m = 3$, which is almost the same as the theoretical calculation. At the same time, the classification time with joint and separate clustering is approximately equal.

Apparently, introducing data clustering into the classification process has significantly reduced computational costs at the expense of a slight decrease in accuracy.

It should be noted that all experimentally obtained accuracy rates directly depend on the selected image database and on the set of alternative images from outside the database, which may vary. The key factor is the visual similarity of the analyzed images to each other, which is directly reflected in the proximity of the descriptor sets. At the same time, a significant gain in processing time and accuracy when using clustering compared to other processing schemes is a significant and objective factor.

VII. CONCLUSION

The traditional method of structural classification is based on a voting procedure and, in practice, implements the nearest neighbor method on a set of descriptors for the etalon description database. The granulation and discarding of features for the etalon descriptions successfully reduce the search volume, which generally increases the classification performance in proportion to the degree of reduction.

Although the choice of the most effective way to compress the description in classification methods depends heavily on the defined set of etalons and the permissible variety of input images, the implementation of intelligent clustering methods has significant advantages over simple data reduction schemes.

Given that the compression of etalon information is performed only once at the preprocessing stage and therefore does not increase the classification time, intelligent approaches, through deeper data analysis, not only simplify calculations but also provide a decent level of accuracy. In addition, integrated data centers are created, which may be practical to implement in many classification cases.

The effectiveness of simplified approaches is random, varies widely, and depends not only on the order of descriptors in the description but also on the implemented processing scheme, which may require multiple checks or selection of different rejection options.

The use of the clustering apparatus has a much higher potential in terms of better accuracy and speed for the classification process compared to simple schemes of screening or granulation by identifying close description components.

ACKNOWLEDGMENT

This work was supported by the European Union NextGenerationEU through the Recovery and Resilience Plan for Slovakia under Project 09I03-03-V01-00115. The results of this research were obtained under the international research project INITIATE under the grant No. 101136775-HORIZON-WIDERA-2023-ACCESS-03.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present research.

REFERENCES

[1] Y. Huang, Z. Cheng, Q. Zhou, Y. Xiang, and R. Zhao, "Data mining algorithm for cloud network information based on artificial intelligence decision mechanism," *IEEE Access*, vol. 8, pp. 53394–53407, 2020, doi: 10.1109/ACCESS.2020.2981632.

[2] H. Li, F. Xie, J. Zhou, and J. Liu, "Object detection, segmentation and categorization in artificial intelligence," *Electronics*, vol. 13, no. 13, p. 2650, Jul. 2024, doi: 10.3390/electronics13132650.

[3] V. Gorokhovatskyi, I. Tvoroshenko, O. Yakovleva, M. Hudáková, and O. Gorokhovatskyi, "Application a committee of Kohonen neural networks to training of image classifier based on description of descriptors set," *IEEE Access*, vol. 12, pp. 73376–73385, 2024, doi: 10.1109/ACCESS.2024.3404371.

[4] Y. Fang and L. Liu, "Angular quantization online hashing for image retrieval," *IEEE Access*, vol. 10, pp. 72577–72589, 2022, doi: 10.1109/ACCESS.2021.3095367.

[5] V. Gorokhovatskyi, I. Tvoroshenko, and O. Yakovleva, "Transforming image descriptions as a set of descriptors to construct classification features," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, p. 113, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp113-125.

[6] R. Tymchyshyn, O. Volkov, O. Gospodarchuk, and Y. Bogachuk, "Modern approaches to computer vision," *Control Syst. Comput.*, vol. 278, no. 6, pp. 46–73, Dec. 2018, doi: 10.15407/usim.2018.06.046.

[7] S. Mashtalir and V. Mashtalir, "Spatio-temporal video segmentation," in *Advances in Spatio-Temporal Segmentation of Visual Data* (Studies in Computational Intelligence), vol. 876. Cham, Germany: Springer, 2020, pp. 161–210.

[8] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, and M. Zeghid, "Improving the effectiveness of image classification structural methods by compressing the description according to the information content criterion," *Comput., Mater. Continua*, vol. 80, no. 2, pp. 3085–3106, Aug. 2024, doi: 10.32604/cmc.2024.051709.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/b:visi.0000029664.99615.94.

[10] X. Zhang, F. X. Yu, S. Karaman, and S.-F. Chang, "Learning discriminative and transformation covariant local feature detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4923–4931, doi: 10.1109/CVPR.2017.523.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.

[12] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555, doi: 10.1109/ICCV.2011.6126542.

[13] V. Gorokhovatskyi, I. S. Tvoroshenko, and O. Đ. Peredrii, "Image classification method modification based on model of logic processing of bit description weights vector," *Telecommun. Radio Eng.*, vol. 79, no. 1, pp. 59–69, Jan. 2020, doi: 10.1615/telecomradeng.v79.i1.60.

[14] D. Padilla Carrasco, H. A. Rashwan, M. Á. García, and D. Puig, "T-YOLO: Tiny vehicle detection based on YOLO and multi-scale convolutional neural networks," *IEEE Access*, vol. 11, pp. 22430–22440, 2023, doi: 10.1109/ACCESS.2021.3137638.

[15] R. Chen, L. Pan, C. Li, Y. Zhou, A. Chen, and E. Beckman, "An improved deep fusion CNN for image recognition," *Comput., Mater. Continua*, vol. 65, no. 2, pp. 1691–1706, 2020.

[16] M. Ghahremani, Y. Liu, and B. Tiddeman, "FFD: Fast feature detector," *IEEE Trans. Image Process.*, vol. 30, pp. 1153–1168, 2021, doi: 10.1109/TIP.2020.3042057.

[17] S. Haykin, "Kernel methods and radial-basis function networks," in *Neural Networks and Learning Machines*, 3rd ed., Upper Saddle River, NJ, USA: Prentice-Hall, 2009, pp. 230–267.

[18] Y. Parzhin, "Principles of modal and vector theory of formal intelligence systems," 2013, *arXiv:1302.1334*.

[19] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, S. Gadetska, and M. Al-Dhaifallah, "Statistical data analysis models for determining the relevance of structural image descriptions," *IEEE Access*, vol. 11, pp. 126938–126949, 2023, doi: 10.1109/ACCESS.2023.3332291.

[20] W. Zhang, G. Chen, P. Zhuang, W. Zhao, and L. Zhou, "CATNet: Cascaded attention transformer network for marine species image classification," *Expert Syst. Appl.*, vol. 256, Dec. 2024, Art. no. 124932, doi: 10.1016/j.eswa.2024.124932.

[21] V. A. Gorokhovatskiy, "Compression of descriptions in the structural image recognition," *Telecommun. Radio Eng.*, vol. 70, no. 15, pp. 1363–1371, Oct. 2011, doi: 10.1615/telecomradeng.v70.i15.60.

[22] P. Flach, "Model ensembles," in *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 330–342.

[23] C. C. Aggarwal, "Advanced topics in neural networks," in *Neural Networks and Deep Learning*. Cham, Switzerland: Springer, 2023, doi: 10.1007/978-3-031-29642-0.

[24] T. Yu, W. Chen, G. Junfeng, and H. Poxi, "Intelligent detection method of forgings defects detection based on improved EfficientNet and memetic algorithm," *IEEE Access*, vol. 10, pp. 79553–79563, 2022, doi: 10.1109/ACCESS.2022.3193676.

[25] Q. Bai, S. Li, J. Yang, Q. Song, Z. Li, and X. Zhang, "Object detection recognition and robot grasping based on machine learning: A survey," *IEEE Access*, vol. 8, pp. 181855–181879, 2020, doi: 10.1109/ACCESS.2020.3028740.

[26] L. Baldini, A. Martino, and A. Rizzi, "Stochastic information granules extraction for graph embedding and classification," in *Proc. 11th Int. Joint Conf. Comput. Intell.*, 2019, pp. 391–402, doi: 10.5220/0008149403910402.

[27] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997, doi: 10.1016/s0165-0114(97)00077-8.

[28] H. Xiong and Z. Li, "Clustering validation measures," in *Data Clustering: Algorithms and Application*. Boca Raton, FL, USA: CRC Press, sec. 2, pp. 571–602.

[29] T. N. Mau, Y. Inoguchi, and V.-N. Huynh, "A novel cluster prediction approach based on locality-sensitive hashing for fuzzy clustering of categorical data," *IEEE Access*, vol. 10, pp. 34196–34206, 2022, doi: 10.1109/ACCESS.2022.3162690.

[30] A. Oliinyk, S. Subbotin, V. Lovkin, O. Blagodariov, and T. Zaiko, "The system of criteria for feature informativeness estimation in pattern recognition," *Radio Electron., Comput. Sci., Control*, vol. 10, no. 4, pp. 85–96, Mar. 2018, doi: 10.15588/1607-3274-2017-4-10.

[31] O. Berehulyak and R. Vorobel, "The algebraic model with an asymmetric characteristic of logarithmic transformation," in *Proc. IEEE 15th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, vol. 2, Sep. 2020, pp. 119–122, doi: 10.1109/CSIT49958.2020.9321906.

[32] O. Yakovleva, A. Kovtunenko, V. Liubchenko, V. Honcharenko, and O. Kobylin, "Face detection for video surveillance-based security system," in *Proc. CEUR Workshop*, Apr. 2023, pp. 69–86. [Online]. Available: https://ceur-ws.org/Vol-3403/paper6.pdf

[33] Z. Ullah, L. Qi, E. J. S. Pires, A. Reis, and R. R. Nunes, "A systematic review of computer vision techniques for quality control in end-of-line visual inspection of antenna parts," *Comput., Mater. Continua*, vol. 80, no. 2, pp. 2387–2421, Aug. 2024, doi: 10.32604/cmc.2024.047572.

[34] J. Crowley and O. Riff, "Fast computation of scale normalized Gaussian receptive fields," in *Proc. 4th Int. Conf. Scale-Space*, Jun. 2003, pp. 584–598, doi: 10.1007/3-540-44935-3_41.

[35] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, and M. Zeghid, "Cluster representation of the structural description of images for effective classification," *Comput., Mater. Continua*, vol. 73, no. 3, pp. 6069–6084, Jul. 2022, doi: 10.32604/cmc.2022.030254.

[36] *ORB Feature Detector and Binary Descriptor*. Accessed: Nov. 22, 2024. [Online]. Available: https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_orb.html
[37] *OpenCV Open Source Computer Vision*. Accessed: Nov. 10, 2024. [Online]. Available: https://docs.opencv.org/master/
[38] *WikiArt*. Accessed: Nov. 26, 2024. [Online]. Available: https://www.wikiart.org/en/maria-primachenko



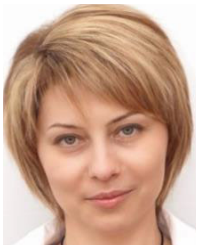
VOLODYMYR GOROKHOVATSKYI received the M.Sc. degree in applied mathematics and engineering from Kharkiv National University of Radio Electronics (KNURE), the Ph.D. degree in management (technical systems), in 1984, and the Dr.Sc. degree in systems and means of artificial intelligence, in 2010.

He received an Internship from Dresden Technical University. He is currently a Professor with the Department of Informatics, KNURE. He has more than 200 scientific articles and seven monographs. His research interests include the image and pattern recognition in computer vision systems, structural methods of image classification and recognition, multidimensional data analysis, and artificial intelligence.



IRYNA TVOROSHENKO received the Ph.D. degree in systems and means of artificial intelligence, in 2010.

She is currently an Associate Professor with the Department of Informatics, Kharkiv National University of Radio Electronics. She has published 182 scientific articles and educational and methodical articles include four study guides, seven monographs, 48 articles, 96 abstracts of reports, 14 lecture notes, and 13 methodological instructive regulations. She is fluent in modern programming languages and technologies, computer-aided mathematical modeling, and constantly expanding her range of scientific interests. Her research interests include image and pattern recognition in computer vision systems, structural methods of image classification and recognition, and fuzzy methods in artificial intelligence appliance.



OLENA YAKOVLEVA received the Ph.D. degree in systems and means of artificial intelligence, in 2004.

In 2008, she obtained the academic title of an Associate Professor with the Department of Informatics. Since 2004, she has been teaching courses on databases, big data, data visualization, artificial intelligence, computer vision, and information systems development with Kharkiv National University of Radio Electronics. Since 2017, she has been a Scientific Consultant in computer vision, machine learning, and intelligent systems with SYTOSS Ltd., Slovakia. Since 2023, she has also been a Researcher with Bratislava University of Economics and Management. She is the author of numerous publications on computer vision, natural language processing, and intelligent systems. Her research interests include computer vision, machine learning, and intelligent systems and related areas.



MONIKA HUDÁKOVÁ received the Ph.D. degree in branch and cross-sectional economics from the Faculty of Operational Economics, Slovak University of Agriculture in Nitra, Slovakia, in 2000, and the Professor degree in economics and business management from Slovak University of Agriculture, in 2022.

Since 2022, she has been an University Professor with Bratislava University of Economics and Management. In 2022, she held the position of the Vice-Rector of Science and Research. Since 2023, she has been the Vice-Rector of Teaching and Lifelong Learning. She is currently a full-time Professor with the Department of Economics and Finance, where she constantly raises the visibility of issues in the field of economics, management, crisis management, and international management. She is also a full member of Czech Academy of Agricultural Sciences and Slovak Academy of Agricultural Sciences. She has received several awards for her long-standing international cooperation and development of science and research. Scientific research activity is documented by a rich, internationally accepted publication activity and participation in numerous research projects.