Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.051709

ARTICLE

Improving the Effectiveness of Image Classification Structural Methods by Compressing the Description According to the Information Content Criterion

Yousef Ibrahim Daradkeh[1,*], Volodymyr Gorokhovatskyi[2], Iryna Tvoroshenko[2,*] and Medien Zeghid[1,3]

[1]Department of Computer Engineering and Information, College of Engineering in Wadi Alddawasir, Prince Sattam bin Abdulaziz University, Al-Kharj, 16273, Saudi Arabia

[2]Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine

[3]Electronics and Micro-Electronics Laboratory, Faculty of Sciences, University of Monastir, Monastir, 5000, Tunisia

*Corresponding Authors: Yousef Ibrahim Daradkeh. Email: y.daradkeh@psau.edu.sa; Iryna Tvoroshenko. Email: iryna.tvoroshenko@nure.ua

Received: 13 March 2024 Accepted: 12 July 2024 Published: 15 August 2024

ABSTRACT

The research aims to improve the performance of image recognition methods based on a description in the form of a set of keypoint descriptors. The main focus is on increasing the speed of establishing the relevance of object and etalon descriptions while maintaining the required level of classification efficiency. The class to be recognized is represented by an infinite set of images obtained from the etalon by applying arbitrary geometric transformations. It is proposed to reduce the descriptions for the etalon database by selecting the most significant descriptor components according to the information content criterion. The informativeness of an etalon descriptor is estimated by the difference of the closest distances to its own and other descriptions. The developed method determines the relevance of the full description of the recognized object with the reduced description of the etalons. Several practical models of the classifier with different options for establishing the correspondence between object descriptors and etalons are considered. The results of the experimental modeling of the proposed methods for a database including images of museum jewelry are presented. The test sample is formed as a set of images from the etalon database and out of the database with the application of geometric transformations of scale and rotation in the field of view. The practical problems of determining the threshold for the number of votes, based on which a classification decision is made, have been researched. Modeling has revealed the practical possibility of tenfold reducing descriptions with full preservation of classification accuracy. Reducing the descriptions by twenty times in the experiment leads to slightly decreased accuracy. The speed of the analysis increases in proportion to the degree of reduction. The use of reduction by the informativeness criterion confirmed the possibility of obtaining the most significant subset of features for classification, which guarantees a decent level of accuracy.

KEYWORDS

Description reduction; description relevance; descriptor; image classification; information content; keypoint; processing speed

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

In image classification tasks, which are now quite relevant for computer vision, a feature system is often formed as a set of multidimensional vectors that fully reflect the spatial properties of a visual object for effective analysis [1,2]. For example, a description of a visual object is represented as a finite set of keypoint descriptors of an image [3–5]. A descriptor is a multi-component numerical vector that reflects the characteristics of certain keypoint descriptor neighborhoods in the image and is formed by special filters–detectors [6,7].

For human vision, different parts of an image have different weights in the process of analysis or classification [8–10]. Human vision, based on the use of mental activity, can recognize objects even by their small details. Even though artificial computer vision systems are now quite effective at recognizing images based on slightly different principles than human vision, the classification importance of the components of an object's image also differs significantly and can be effectively used in the analysis process.

Factors that characterize the importance (weight, significance, significance) of features: the spread of values within the description and the database of etalon descriptions, the degree of resistance to geometric transformations and interference, the uniqueness and value of the presence in the image, etc. However, the specific level of the significance parameter, as well as the classification performance in general, is largely determined by the base of etalon images to be classified. The introduction of the significance parameter in the process of classification analysis allows us to move from the homogeneous influence of features to taking into account their relative weighting, which directly affects such classification characteristics as performance and efficiency [3,11,12]. As a rule, the significance of individual image features in the classification process is taken into account by using weighting coefficients in numerical terms. The significance can be estimated a priori for the available set of components of the etalon descriptions, and its use is aimed at better adaptation of the classification to the analyzed data due to the expansion of the amount of information [2,13]. The effectiveness of the implementation of significance parameters is also influenced by the form of representation of the feature space and the classification method.

If you first analyze the calculated weight values for the etalon components of the image database descriptions, you can construct class descriptions from the most valuable elements for classification, discarding the uninformative part of the description [2,14,15]. One of the effective estimates of classification significance is the feature informativeness parameter [2]. The higher the informativeness of a feature, the better this feature divides the instances of the training set into classes. This not only reduces computational costs in proportion to the degree of reduction but also preserves and often improves the efficiency of classification systems. The introduction of weighting coefficients, in particular, the informativeness parameter, makes it possible to perform a deeper data analysis, which generally improves recognition accuracy.

The work aims to reduce the amount of computational costs when implementing structural methods of image classification while maintaining their effectiveness. Reducing the description in the form of a set of keypoint descriptors is achieved by forming a subset of descriptors according to the criterion of classification significance.

The objectives of the research are as follows:

1. Studying the influence of the informativeness parameter for the set of descriptors of etalon descriptions on the effectiveness of image classification.
2. To reduce the etalon descriptions based on the value of informativeness.

3. Implementation of the informativeness parameter in the classifier model.
4. An experimental research of the effectiveness of the developed modifications of classifiers in terms of accuracy and processing speed based on simulation modeling for an applied image database.

2 Formal Problem Statement

Let us consider the set $E = \{E_i\}_{i=1}^N$ of descriptions of the etalon database $E = \left\{\{e_v(i)\}_{v=1}^s\right\}_{i=1}^N$, where N is the number of classes, $i = 1, \ldots, N$ is the class number, s is the power of the set of components of a separate description, $v = 1, \ldots, s$ is the number of the component within the class description E_i. Thus, the data model for the classification database E can be represented in the form of a matrix of size $N \times s$, where each row contains a description of a separate etalon (class) as part of s the components–keypoint descriptors [15].

The description $Z = \{z_v\}_{v=1}^s$ of the recognized object is represented by a model that similar to the description of a separate etalon E_i. The latest keypoint detectors use binary representations for keypoint descriptors, which gives significant advantages in terms of computation [1,6]. We assume that $z_v \in B^n$, $e_v \in B^n$, $E_i \subset B^n$, $Z \subset B^n$, where B^n is a space of vectors of dimension n with binary components.

The finite set of binary vectors (descriptors) obtained by the keypoint detector creates a transformation-invariant description of the etalon or recognized image [16]. An etalon is a selected image for which a description is generated. A set of etalons creates a basis for classification. Formally, the recognized class in this formulation [1,4] has the form of an infinite set of images obtained from the image of a particular etalon by applying to it various sets of geometric transformations of shift, scale, and rotation.

For each element $e_v(i) \in E$ of the etalon database, at the preparatory stage of data analysis, we will determine the value of the significance parameter $\lambda_{i,v}$ in the form of a number. In general, for the etalon database E, we will obtain the matrix $\Lambda = \left\{\left\{\lambda_{i,v}\right\}_{v=1}^s\right\}_{i=1}^N$ of numbers

$$\Lambda = \begin{Bmatrix} \lambda_{1,1}, \ \ldots, \lambda_{1,s} \\ \lambda_{2,1}, \ \ldots, \lambda_{2,s} \\ \ldots \\ \lambda_{N,1}, \ \ldots, \lambda_{N,s} \end{Bmatrix}. \tag{1}$$

The matrices Λ and E have an identical structure. However, the components of E are binary vectors, and the components of Λ are numbers that characterize the significance of these vectors for classification. The rows of the matrix E contain the values of keypoint descriptors for the descriptions of individual classes; Λ contains a quantitative parameter of significance for each of the descriptors.

It should be noted that the parameter $\lambda_{i,v}$ of the matrix Λ, as a rule, reflects the weight of each feature $e_v(i)$ within a separate class E_i or the entire available set E of classes. At the same time, the known methods [2,11,13] for determining the elements of $\lambda_{i,v}$ indicate their direct dependence on the composition of E, since the classification weight of the components directly and always depends on the set of recognized classes.

Let us set the task of developing a procedure $R: E \rightarrow E^*$ for a targeted reduction of the composition of the description of the database E by selecting the most significant elements according to the criterion of the matrix Λ, and $card\ E^* << card\ E, E^* = \left\{E_i^*\right\}_{i=1}^N$. The power $s^* = card\ E_i^*$ of the description of each of the etalons E_i is equally reduced: $s \rightarrow s^*, s^* << s$. The formation of the newly created database E^* based on the reduction of R is intended to perform classification with significantly lower computational costs and ensure a decent level of performance.

The second urgent task is to create a classifier F, which, for an arbitrary description $Z = \{z_v\}_{v=1}^{s}$ of an object in the form of a set of keypoint descriptors, will make a decision $F[Z] \to [1, 2, \ldots, N]$ about whether the analyzed description Z belongs to one of N classes, taking into account the values of the matrix Λ. In some situations, it is possible to refuse classification.

In the formulation under discussion, the classification of $F[Z]$ is based on the following principles: descriptions Z of the object and the database $\{E_i\}_{i=1}^{N}$, as well as the determined significance in the form of a matrix Λ for the set of etalon components. The full formal parameterization of the classifier is in the form of $F[Z, E, \Lambda] \to [1, 2, \ldots, N]$. At the same time, in the classifier F, it is allowed to use the reduced database E^* instead of E. This requirement needs to be modified for the model to assess the relevance of the descriptions Z and E_i^* with an unequal number of elements.

3 Literature Review

It is worth noting that keypoint descriptors are a modern and effective tool for representing and analyzing descriptions of visual objects [6,7]. This technique effectively ensures invariance to common geometric transformations of visual objects in images, high speed of data analysis, and decent classification performance [1,4]. Unlike neural network models that generalize an image within a class, these methods focus on identifying the characteristic properties of recognized images, which is necessary for several applications [13,17–20].

Improving the performance of structural classification methods based on comparing sets of vectors is being developed in such aspects as speeding up the process of finding component correspondences by clustering or hashing data [4,17,21–24], using evaluation to narrow the search scope [25], forming aggregated features in the form of distributions [4,26], and by determining the most important subset of features for classification [2,14].

The task of forming an effective subset of features, including by reduction, is constantly in the field of attention of computer vision researchers [13,17,18] due to the need to ensure productive analysis and processing of large multidimensional data sets in such systems [27]. Papers [2,3,15] consider the determination of the informativeness parameter with respect to the training set's features, and paper [15] studied the features of calculating the informativeness of descriptor vectors in more detail. Papers [4,9,11] discuss ways to determine features' significance based on the description's principle of uniqueness and the form of a significance vector for a set of classes.

The data analysis literature studies some models for using feature significance to improve classification accuracy [13,19]. To do this, when comparing the analyzed vector numerical description $\{a_v\}_{v=1}^{s}$ with the etalon $\{b_v\}_{v=1}^{s}$, if their components are consistent, a certain measure $\mu(a, b, \lambda)$ is determined, taking into account the significance parameters $\lambda = \{\lambda_v\}_{v=1}^{s}$ of each feature. Such a measure takes the form of a weighted metric, for example, the Manhattan metric:

$$\mu(a, b, \lambda) = \sum\nolimits_{v=1}^{s} \lambda_v |a_v - b_v|, \sum\nolimits_{v=1}^{s} \lambda_v = 1. \tag{2}$$

In measure Eq. (2), the significance of λ_v is used in conjunction with the metric, meaning that more important features contribute more to the classification decision. Given that the metric in Eq. (2) is set to search for the minimum when the descriptors of a, b are matched, the vector λ should be adapted to this classification method. The λ parameter should increase the similarity measure for more important features. By definition, it is larger for more important features. Given the existing condition $\sum_{v=1}^{s} \lambda_v = 1$, one of the variants of the model Eq. (2) is to use the value $(1 - \lambda_v)$ instead of λ_v. Note that when directly applying Eq. (2) according to the linear search scheme "each with each" for descriptions as sets of vectors, it is often necessary to operate with floating point numbers, which requires some increase in the amount of computation or discretization of values λ [1,13].

The weighting coefficients for keypoint descriptors have been used in several probabilistic classification models, where these coefficients are interpreted as the probability of classification [4,18,28]. According to the developed schemes, the values of those etalon elements are accumulated for which the correspondence with the analyzed component of the object is established competitively [19]. The introduction of the model Eq. (2), according to Biagio et al. [19], contributed to a better adaptation of the data analysis process compared to a fixed grid in the space of classification features. It is worth noting that the weighting apparatus is also successfully used in neural network models for data classification [29]. There, they are usually used for classification as a result of network training [30,31].

Today, the applied performance of modern neural network systems [30,32] in the task of image classification is very difficult to surpass. However, these systems are known to require long-term training on large sets of data already partially annotated by humans using specialized software. At the same time, the result and efficiency of classification by neural networks are significantly determined by their structure and the composition of the data used for training. In addition, neural networks traditionally generalize features for representatives within a class [31], which does not always make it possible to perform an effective classification for its members.

Approaches based on the direct measurement of image features in the form of a set of descriptors have their advantages when implemented in computer vision systems [4,6,7,33–35]. Their positive aspects are the absence of the need for a training phase, as well as the possibility of rapid and radical changes in the composition of recognized classes. They can be most effectively implemented for identifying or classifying standardized images (coats of arms, paintings, brands, museum exhibits) with permissible arbitrary geometric transformations of objects in the field of view [3,21,33]. For the functioning of such methods, only representatives of classes–etalons–are required. After a short time of calculating the descriptions of the etalons, the method is ready for use. The methods are universally suitable for any selected set of etalons, the composition of which can be quickly changed for an applied task [1].

It is clear that the keypoint descriptor apparatus does not have the ability to take into account the almost infinite variants of generative models for transformations in the formation of images by modern neural networks [36]. Therefore, the effectiveness of its application for classifying such a variety of images should be studied separately.

It is possible that these research areas (neural networks and structural methods) should be divided by application areas or applied tasks. For example, for neural networks, it is difficult to cope with the variability of objects in terms of geometric transformations. Also structural methods are sensitive to significant changes in the shape of objects (morphing), and when tracking moving objects, they require rewriting the etalon.

There is known research [33] on determining the number of descriptors in the description and selecting a keypoint detector with the best recognition performance. However, in this formulation, the emphasis on the result and efficiency of classification is focused solely on the type and properties of the detector [37], not on the form of data and the method of its classification.

Thus, the conducted research indicates the need for a more detailed study of the process of implementing classification weighting indicators and evaluating their impact on the effectiveness of classification by description in the form of a set of descriptors both based on reducing the set of features and by direct use in classifiers in conjunction with metric relations.

4 Materials and Methods

4.1 Implementation of Significance Parameters for Description Components

An important factor that can affect the classification result can be considered the use of individual values of significance Eq. (1) for individual components of the etalon description [4]. Given the fact that a classification decision is usually based on statistical analysis for a set of values of metric relations in the descriptor space, we will enhance it by introducing a significance parameter. This expands the set of parameters for the data analyzed by the classifier.

To do this, let's consider one of the practical schemes based on the initial classification of individual object descriptors with the subsequent accumulation of certain votes for their classes and values [1,4].

Considering that the order of components in the descriptions of the etalon and the recognized object Z may differ due to the possible influence of geometric transformations of the object, we apply a two-stage classification scheme, where at the first stage we classify $z_v \rightarrow [1, 2, \ldots, N]$ for each component $z_v \in Z$. This scheme compares favorably with the integral representation of objects [21,26] in that it can reject false background and interference descriptors, which provides better noise immunity [11].

According to the traditional nearest-neighbor scheme, we will classify by calculating the minimum

$$k = \arg \min_{v,i} \rho \left(z, e_v \left(i\right)\right), \tag{3}$$

where $\rho\left(z, e_v\left(i\right)\right)$ is the distance between the object and etalon descriptors, $k = 1, \ldots, N$ is the determined class for the descriptor z, which corresponds to the class of the etalon element with the smallest distance. The process of determining the class k in Eq. (3) is a two-parameter optimization with the parameters of the number i of the class and the number v of the etalon descriptor within the class.

The main metric used to evaluate the deviation of a pair of binary descriptors is the computationally efficient Hamming distance [2,17], which determines the number of distinct bits.

The model for implementing value largely depends on the chosen method of implementation Eq. (3). By the traditional model Eq. (3), for each descriptor $z \in Z$, we find the minimum ρ_m distance $\rho_m = \min_{v,i} \rho\left(z, e_v\left(i\right)\right)$ on the etalon set of the database E and determine its class number k. An important condition for taking into account the class and significance is the fulfillment of the inequality $\rho_m \leq \delta_z$, where δ_z is a boundary parameter that a priori determines the significance of the minimum metric for establishing an equivalence between descriptors [1].

After performing Eq. (3) for an individual $z \in Z$, we increment the accumulator $h = (h_i)_{i=1}^{N}$ as $h_k = h_k + 1$ for the resulting class number and aggregate the significance vector $val = (val_i)_{i=1}^{N}$ as $val_k = val_k + \lambda_{k,v}$ for this class. The value of $\lambda_{k,v}$ is the significance of the representative of the etalon with the number k, for which the minimum in Eq. (3) is achieved. As a result of processing the entire description Z of the object, we obtain the accumulated vectors h, val, which can form the basis of the classification decision.

As a result of the first stage, the final number of h_i votes and weights val_i are obtained based on those subsets of the object's components that are assigned to each of the classes E_i, $i = 1, \ldots, N$. The vector $\{val_i\}_{i=1}^{N}$ is formed as the sum of the numbers $\lambda_{i,v}$ for those components $e_v\left(i\right)$ for which the correspondence to the elements $z_v \in Z$ is established according to the procedure Eq. (3). In this scheme, the metric relation remains fundamental to the classification process, and the significance plays a secondary role. This can be explained by the fact that the descriptor is the main feature of the image, and the significance of the $\lambda_{i,v}$ descriptor is a function of both the content of the etalon database and the way it is calculated.

The final classification decision F on the class k for the object Z has the form

$$k = \arg \underset{i=1,\dots,N}{opt} F\left[\{h_i\}_{i=1}^{N}, \{val_i\}_{i=1}^{N}\right]. \tag{4}$$

Model Eq. (4) is a development and generalization of the voting scheme using the weighting coefficients of the matrix Λ and requires detailed study and specification for a variety of options F.

Let us define the possible practical conditions for refusing to define a class as

$$\max_i \{h_i\} < \delta_h, \ \max_i \{val_i\} < \delta_{val}. \tag{5}$$

Condition Eq. (5) corresponds to a situation where the accumulated maximums among the number of votes and/or values are quite low, which makes it impossible to confidently determine the class. It is clear that the thresholds δ_h, δ_{val} are estimated experimentally and directly depend on both the content of the etalon database and the implementation scheme Eq. (3).

Along with the conditions Eq. (5) that set absolute constraints, application systems also use relative constraints on the significance of the maxima Eq. (5) compared to the closest competitor for another class [1]. Such constraints additionally contribute to the confidence of the classification solution and significantly affect the result. The significance matrix Λ is calculated at the preparatory stage and insignificantly increases the computational cost of the classification process. Given that image descriptors are formed in an arbitrary order, each vector row of the Λ matrix should be considered as a set of values in the analysis process.

As shown by the results of our experimental modeling using specific models Eq. (4) (experiments section), among the possible variants of Eq. (4) for constructing an optimal two-parameter solution, the simpler ones that optimize the number of votes are more practical and effective $\{h_i\}_{i=1}^{N}$. The use of significance in the form of informativeness [2] correlates significantly with the number of votes and can be used to confirm the decision.

4.2 Methods of Counting Class Votes

By the traditional scheme Eq. (3) for determining the nearest neighbor, distance optimization is carried out using two parameters–the class number and the number of the descriptor within the class. In the literal sense, scheme Eq. (3) implements an aggregate decision by an ensemble of simple classifiers for the aggregate set of object descriptors [4]. The number of effective votes used to make a decision in this scheme is equal to the power of the analyzed description Z.

It is clear that in classification, the main result is the class number. Therefore, in practice, especially when the etalons are set, more productive approaches are often used, which are reduced to a stepwise search, where the first step is a search within the description for a fixed class [21]. The basis of such approaches is the fundamental fact that any classification procedure based on a description in the form of a set of vectors in the most general aspect evaluates the degree of relevance of the object–etalon, which is optimized on a set of etalons. In turn, such a relevance measure directly reflects the intersection power of two finite sets of vectors (object and etalon descriptions). In addition, stepwise approaches are based on important a priori information about the etalon descriptors belonging to a fixed class, which makes the classification more robust by using consistency with the etalon data.

It should be noted that the implementation of each of these modifications is associated with some peculiarities (the method of determining the relevance of descriptors, the choice of the threshold for the number of votes, significance, etc.) In addition, if the number of descriptors in a description is small or if the powers of the compared descriptions are unequal, there are additional difficulties with the "friend-or-foe" distinction [25,33].

As an option, let us consider a classification method that consists of accumulating optimal values for each of the classes without first setting the class number for a particular descriptor, as is done in the traditional method Eq. (3). This is one of the practical options for implementing Eq. (3). It can be considered a modification of the nearest neighbor method.

We find the minimum distances

$$\rho_{m,v}(i) = \min_{z} \rho\left(e_v(i), z\right), \; i = \overline{1, N}, \; v = \overline{1, s}, \tag{6}$$

separately for each descriptor $e_v(i)$ of the class E_i among the elements Z. If the condition $\rho_{m,v} \leq \delta_z$ holds, we accumulate the vectors of votes $\{h_i\}_{i=1}^{N}$ and values $\{val_i\}_{i=1}^{N}$ according to the additive model $val_i = val_i + \lambda_{i,v}$, where $\lambda_{i,v}$ is the value of the descriptor from the class i, for which the optimum $\rho_{m,v}(i)$ is achieved. The decision about the class is made according to Eq. (4) based on the accumulated votes and values.

This method is more focused on consistency with etalon data, when each etalon "looks for its own" among the components of the object description. This method is effective when the power of the etalon and object descriptions is different. The maximum number of received votes for a class coincides with the number of descriptors in the etalon.

Another practical approach to establishing the degree of relevance of two descriptions is to search for a minimum with double checking (Cross-Checking [38]): for the object descriptor that corresponds to the found minimum, the minimum distance among the etalon descriptors is counter-determined; the correspondence of the object and etalon descriptors is established only if the result for both searches is the same. Such double-checking can be performed for matching procedures both within the entire database and separately for each of the etalons. The introduction of Cross-Checking models is aimed at increasing the reliability of classification decision-making by reducing the number of possible outliers.

In general, the data analysis scheme by models Eqs. (3), (4) and (6) can be enhanced by identifying not one but several nearest minima, for example, three [4,13]. This corresponds to the "three nearest neighbors" model. In this way, at the second stage of classification, three classes will receive the number of votes and the aggregation of values (they can be the same), and the decision-making scheme Eq. (4) will not change.

In the considered models, the class of the analyzed descriptor, taking into account the significance parameter, is determined based on two criteria: the value ρ_m of the minimum distance to the defined etalon descriptor and the a priori set of significances for the etalon descriptors obtained during the training phase. Note that the distance estimation in Eqs. (3) and (6) can also be performed using modern high-speed methods based on the data hierarchy instead of the linear search, which is quite computationally intensive [1,25,38].

4.3 Performing Description Reduction

One of the most effective ways to improve classification performance is to compress the feature space [2,39]. For classifiers based on a description as a set of keypoint descriptors, this method can be implemented by reducing the power of the etalon descriptions. In this case, the classification speed increases in proportion to the degree of reduction. The value of the matrix Λ can be used as a criterion for selecting significant components of the etalon descriptions. Based on the a priori analysis of the matrix Λ, the volumes of the descriptions of the etalon E_i can be significantly reduced because their components with low significance are excluded from the classification process. It should be noted that after the description is shortened, the values of the significance change themselves, as the composition of the database changes. Therefore, the value Λ must be recalculated if a modified description is used in the classification process with the introduction of values [2]. According to our experimental research, we can recommend a scheme where the description is reduced in stages, and at each stage, the coefficients for the reduced composition of the rows of the matrix Λ are recalculated.

Given the peculiarities of the voting method, where the number of comparable descriptions is considered to be approximately the same, it is natural to consider the dimensions of the descriptions transformed after compression to be equivalent [40]. However, a more detailed study reveals significant differences between the powers of the object descriptions and the etalons, which requires modification of the procedure for counting and analyzing votes.

Our modification of the nearest neighbor method using the Cross-Checking model makes it possible to realize this. As a result of the reduction of the etalon descriptions, it becomes necessary to modify the parameters of the classifier: the decisive number of votes, the ratio of global and local maxima, etc.

Our analysis has shown that direct selection of a fixed reduced number of keypoint descriptors by controlling the keypoint detector parameters does not lead to improved performance, as the classification accuracy decreases in proportion to the description reduction. It is more effective to initially generate a large number of description descriptors (500 or more) that reflect all image features, followed by a reduction based on the significance criterion.

It should be noted that if the number of descriptors for the etalons and the input image differs significantly, the probability of a degenerate situation when several etalons can be found in the input image increases somewhat. Thus, each reduction has its limits of application.

The classifier scheme using description reduction is shown in Fig. 1. The green color in Fig. 1 shows the additional data processing blocks introduced by us using the reduction based on the significance of the informativeness parameter [2,15].

4.4 Determining the Significance Parameters

Given that humans often form the recognition conditions in computer vision systems, the value of the matrix Λ can be formed a priori by an expert, and the human eye estimates the significance parameter by selecting the most important keypoint scans for classification. However, to ensure the automatic functioning of the systems, it is necessary to have models for assessing the significance of the data. Note that, all the significance parameters discussed here are calculated on the training set of features.

Let us use the metric criterion of informativeness to calculate the matrix Λ [2]. For an arbitrary vector $z \in E$ in the class system as a component element $z \in E_k$ of a fixed etalon description E_k with the number k, we introduce the concept of informativeness $I(z, E)$ as part of the base E

$$I(z,E) = \rho_m\left(z,\overline{E}_k\right) - \rho_m(z,E_k), \tag{7}$$

where $\rho_m\left(z,\overline{E}_k\right) = \min_{v,i\neq k}\rho(z,e_v(i))$ is the minimum distance from z to an element of the database that does not belong to the class E_k, $\rho_m(z,E_k) = \min_{v,i=k}\rho(z,e_v(i))$ is the distance from z to the nearest element of the class E_k (excluding the distance $\rho(z,z) = 0$ to itself, $z \in E_k$).

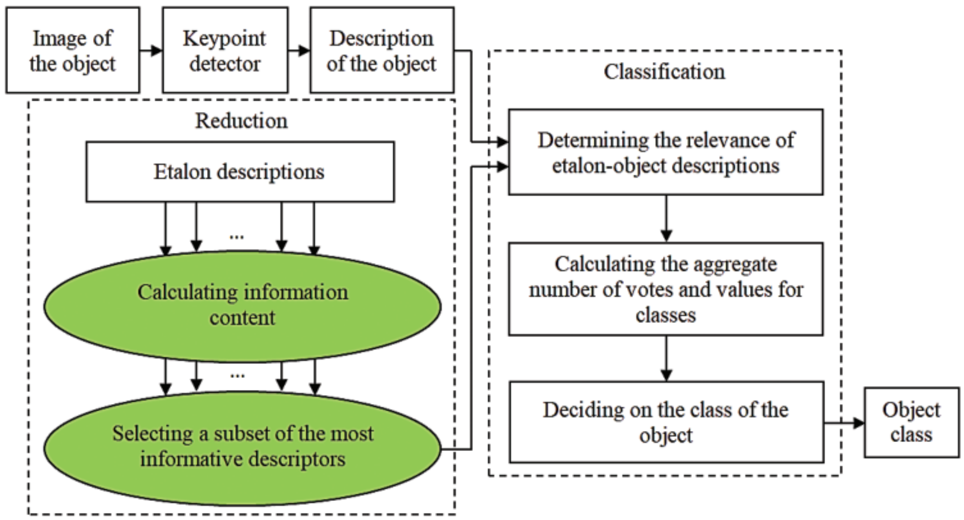


Figure 1: Classification scheme using description reduction

When implementing normalized distances with a value of $0 \le \rho \le 1$, the estimated values of criterion Eq. (7) lie in the interval $-1 \le I \le 1$. The Relief method uses one of the variants of the model Eq. (7), where the informativeness is estimated as the difference of squared distances, and the selection of the analyzed features is carried out randomly [41]. As we can see, criterion Eq. (7) has the property of data evaluation.

The use of model Eq. (7) to determine the individual informativeness of I for $z \in E_k$ is based on the assumption that the classification value of a feature is higher the better it divides the instances of the training set into classes. Based on this, the farther away from an instance of a class is the closest instance of another class, the higher the individual informativeness. At the same time, the farther away from the instance the closest element of the same class is, the lower the individual information content. In other words, the principle "closer to your own, farther from others" is implemented. Thus, features Eq. (7) with high values of I of individual informativeness will be considered significant, i.e., significantly informative concerning the effective classification, features with low I are considered insignificant, i.e., candidates for exclusion from the feature system.

In [11], we consider other criteria for assessing significance, in particular, those based on the principle of uniqueness of a feature among E elements inside and outside the class. Other principles of forming significance criteria based on the stability of descriptors under geometric transformations and using distributions by defined classes, including cluster representation, are considered in [2,4,17,18,21,35].

Criterion Eq. (7) can be considered more effective than the others since it already partially reflects the degree of metric distinction of the analyzed data components, which is the basis for classifying images according to the generated descriptions. In addition, criterion Eq. (7) is characterized by a number, which simplifies processing.

Thus, our proposals for improving structural classification methods [1,5,15] in order to reduce computational costs are as follows:

1) In accordance with the classification scheme of Fig. 1 for the entire set of s descriptors in each etalon, we calculate the informativeness index (7).

2) Based on the largest values of the indicator (7), we select a fixed number of $s^* << s$ descriptors in each etalon.

3) We use the obtained subset of s^* descriptors in the classification scheme (model (6), Fig. 1).

4.5 Performance Criteria and Classification Thresholds

The classification performance will be evaluated by the value of the accuracy index pr, which is calculated by the ratio of the number of correctly classified objects r_p to their total number r, which was used in the experiment [13].

$$pr = r_p/r. \tag{8}$$

Another important indicator for voting methods is the ratio of the maximum number of votes $h_{\max 1}$ or aggregate value, by which the classification decision is made, to the nearest local maximum $h_{\max 2}$ of the competitor class

$$\Delta = h_{\max 2}/h_{\max 1}. \tag{9}$$

The value Δ characterizes the degree of confidence in the decision. The smaller the Δ, the more reliable the decision on the class of the object.

The main thresholds used in the paper are threshold δ_z for the value of the metric for recognizing two descriptors as equivalent and threshold δ_h for the number of descriptor votes required to make a classification decision based on the maximum number of votes. Both thresholds are chosen experimentally in the paper and depend on the selected etalon database.

As for the equivalence threshold δ_z, there are literature sources [13], where a threshold of 25% of the maximum of the metric for multidimensional vectors is considered acceptable. Although, according to our research [14,15], this threshold can be determined in a more adapted and efficient way. In this paper, the threshold δ_h for the number of votes according to the statistical experiment is optimally chosen so that the defined threshold exceeds the minimum among the maximums of the votes of the transformed etalon images, but at the same time is higher than the maximum among the votes of the images outside the database. It is not possible to optimize these thresholds analytically due to the unlimited amount of real images and their descriptors.

5 Experimental Results and Discussion

Note that the value of the informativeness parameter Eq. (7) as the difference between two distances is not directly related to the probability of correct classification, unlike the value of similarity or a metric between data sets [13,28,33]. Therefore, the effectiveness of its use for recognized data should be tested experimentally, which will make it possible to evaluate the applied classification accuracy using informativeness.

First experiment. Using the software tools of the OpenCV library, the Binary Robust Invariant Scalable Keypoints (BRISK) detector generated a description from a set of keypoint descriptors (a BRISK descriptor contains 512 bits) for the analyzed images [6,42]. In the experiment, for three different images of the same fairy-tale character (Fig. 2) with the number of $s = 500$ BRISK descriptors in the descriptions by selecting 100 descriptors with the highest information content Eq. (7), we predicted to achieve about a fivefold gain in classification time. The classification accuracy for the full and reduced descriptions is $pr = 1.0$, i.e., all images are classified correctly.

Figure 2: Image of the reduction experiment according to criterion Eq. (7)

At the same time, the classification confidence index Eq. (9) for the reduced description of $s = 100$ descriptors was 0.29, while for the full description of 500 descriptors, this index for the analyzed images of Fig. 2 is 0.59, which is much worse. The obtained experimental result shows that even for sufficiently similar images, a significant increase in reliability in terms of image distinction has been achieved due to the reduction of the information content coefficient. Reducing the description five times by the informativeness criterion not only gives a significant gain in processing time but also provides an increase in the confidence index while maintaining the accuracy of the classification solution. The selected subset of the most informative descriptors fully ensures the classification performance with significantly lower computational costs.

Second experiment. To study the effectiveness and properties of reduced descriptions for classification in more depth, we conducted a large-scale experiment using a large-scale test material. The classification accuracy is evaluated for different variants of the classifier and a reduced description is presented with a reduction in the number of s descriptors in the description in a wide range from 500 to 10.

For software modeling, the Python programming language, the OpenCV computer vision library, and the NumPy library were used to accelerate the processing of multidimensional data, and 500 256-bit Oriented FAST and Rotated BRIEF (ORB) descriptors were created to describe each image [7,42–44]. Images of jewelry from the National Museum of History of Ukraine were used as the classification base [45]. For testing, a classification base of 5 images of museum exhibits Fig. 3 (diadem (a), pectoral (b), pendant (c), topelik (d), zuluf-ask (e)) was formed and used, as well as the other 3 images of the jewelry collection Fig. 4 (buckle (a), pendant No. 1 (b), pendant No. 2 (c)), which are not included in the classification base.

For testing, a set of input images from Figs. 3 and 4, where each image is additionally transformed by applying geometric transformations for 6 different combinations of rotation and scale. Thus, the test set of 51 images includes 30 transformed images of the base and 21 images of non-base decorations. Fig. 5 shows examples of transformed tiara images (45° rotation, 80% scale (a); 30° rotation, 120% scale (b)).

(a) (b) (c) (d) (e)

Figure 3: Images of the etalons (a) diadem, (b) pectoral, (c) pendant, (d) topelik, (e) zuluf-ask from the classification database

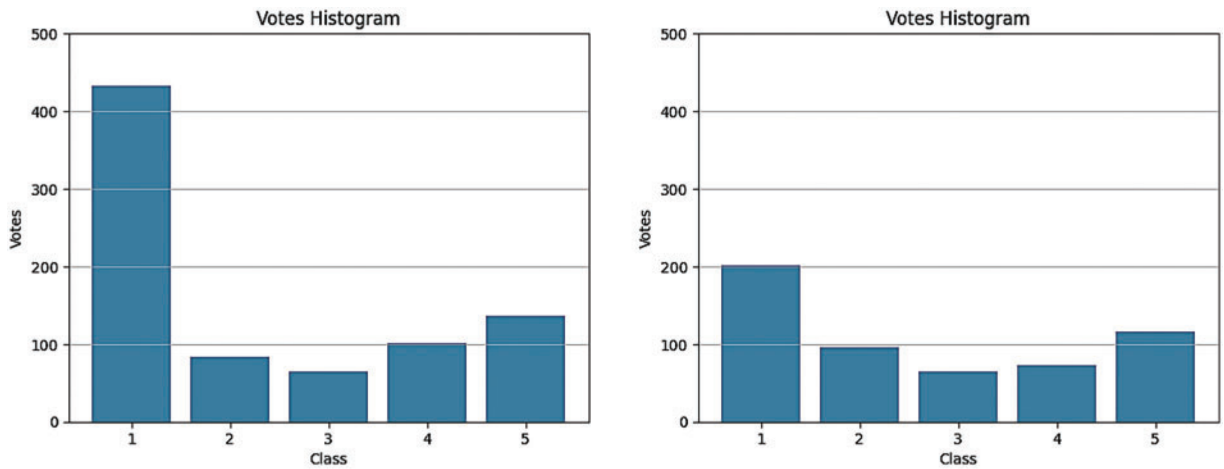
(a) (b) (c)

Figure 4: Test images of jewelry (a) buckle, (b) pendant No. 1, (c) pendant No. 2 not included in the classification database

(a) (b)

Figure 5: Transformed images of the tiara (a) 45° rotation, 80% scale; (b) 30° rotation, 120% scale with coordinates of keypoints

For a given test set with several descriptors of $s = 500$, all 51 images were classified correctly by the traditional method (the transformed etalon images were assigned to their "own" class, and the rest were assigned to none of the classes). At the same time, for the transformed etalon images, the Δ classification confidence did not exceed the value of $\Delta \leq 0.1$, and for images not from the database, there was a uniform distribution of votes for the classes at the value of $\Delta > 0.45$, which made it impossible to classify them. A typical example of the number of votes for a non-database image: [212, 107, 80, 105, 139], coefficient $\Delta = 0.66$ (Fig. 6). An example of a histogram of votes for an existing set of 5 classes for transformed etalon images (a) and out-of-base objects (b) is shown in Fig. 6. As we can see, the histogram for the transformed etalons has its single-modal shape, which facilitates classification, and for the out-of-base object, the votes are distributed quite uniformly.



(a) (b)

Figure 6: Vote histograms for the transformed (a) etalon and (b) the out-of-base object

Despite the lower threshold of Δ for some out-of-base images, none of them received more than half of the votes (250). The obtained indicators for the number of votes generally indicate a significant similarity of all used images (out-of-base and within the database) in the constructed feature space.

Next, based on the calculation of indicator Eq. (7), the 50 most informative descriptors for each of the etalons were identified (Fig. 7a). Thus, a reduced description of the etalons was formed from the 50 selected descriptors (Fig. 7b).

(a) (b)

Figure 7: Coordinates of (a) 500 (red) and (b) reduced 50 (blue) keypoints

Given the unequal number s of keypoint descriptors in the reduced etalon ($s = 50$) and the input image ($s = 500$), a classification method is applied where each etalon selects "its" equivalent descriptors on the object, the maximum possible number of which is now equal to the volume of the etalon description. The resulting minimum Hamming distance is checked according to the condition $\rho_m \leq \delta_z$, $\delta_z = 64$ (the number 64% is 25% of the maximum value of 256 for the Hamming metric [13]), and is counted as a vote for the class when it is performed.

It should be noted that the use of the δ_z threshold to test for the significance of the calculated minimum remains a controversial point and needs to be studied and evaluated for each of the specific classification bases, despite the undisputed statement of respected researchers about the significant similarity of the elements of the multidimensional vector space [13,17].

The value of δ_z, statistically dividing descriptors into "own and others", is leaving a certain tolerance for deviations of "own". As a rule, this threshold is determined based on classification experiments for "own", since "others" belong to the rest of the virtually infinite space B^n of multidimensional binary vectors. In some applications, different values of the δ_z threshold are used for system training or parameterization and separately for the classification process [1,28].

Based on our experiments in this and other research [4,35], we can say that in some situations, classification without using the δ_z threshold is even more effective. Therefore, in applied applications, it is necessary to research all possible options and choose the best method, the choice of which may depend on the composition of the recognized data.

Based on the experiments, the threshold for the effective number of votes was statistically chosen as half of the number of descriptors (250 and 25, respectively). The threshold $\delta_\Delta = 0.62$ for the confidence coefficient Δ was also experimentally selected, which ensures the correct class determination for all images from the database. The classification decision is formalized by the following situation: the number of votes for the winning class exceeds half, and the value of the confidence coefficient is less than 0.62.

Given that any progressive system strives for simplification, the δ_Δ threshold, in our opinion, should not be included in the classifier's operation. This is especially true for experiments with a full description of 500 descriptors. The δ_Δ indicator can be used to a greater extent to control the situation, as it characterizes the degree of reliability of solutions. However, for small amounts of description (50, 25, 10), our analysis has shown that the use of the δ_Δ indicator becomes more relevant, as it makes it possible to separate "foreign" descriptors.

For the test reduced set at $s = 50$, all input images were classified correctly, but the Δ classification confidence score for the transformed etalon images predictably increased to the level of $\Delta \leq 0.6$, and for images not from the database, the distribution of votes for classes was obtained at even higher values of $\Delta > 0.7$. We can see that the Δ indicator in a situation with data reduction makes it possible not only to evaluate but also to ensure the reliability of classification.

The accuracy of pr classification was estimated as the ratio Eq. (8) of the number of correctly classified objects to the total number of experiments. The experimental accuracy rate was 1.0 (maximum) both for descriptions of 500 descriptors and for the reduced base of 50 most informative etalon descriptors, which once again confirms the high efficiency of the method with the introduction of reduced descriptions. At the same time, the classification time for one image decreased by about 10 times from 0.27 s for 500 descriptors to 0.027 s for 50 description descriptors (Table 1).

For comparative evaluation, in our separate experiment, we directly initially determined 50 and 25 (instead of 500) ORB descriptors for the object and the etalons. Such a direct reduction of the analyzed data and its use for classification led to a significant deterioration of the *pr* accuracy rate for 50 descriptors to the level of 0.6...0.7, and for a description of 25 descriptors–even to the level of 0.4...0.5. The decrease in the accuracy of *pr* here was more influenced by the test images that were not included in the classification database.

Table 1: Experimental performance indicators of classification methods

Parameters	Threshold δ_h for the number of votes				Time t_c classification, s				Accuracy *pr*			
Number of descriptors	500	50	25	10	500	50	25	10	500	50	25	10
Modification of the nearest neighbor	250	25	15	8	0.28	0.027	0.014	0.0061	1	1	0.95	0.93
Cross-Checking	220	25	15	6	0.28	0.026	0.013	0.0056	1	1	1	0.96

Thus, only the method of description reduction based on determining the informativeness of the etalon descriptors maintained a high level of classification accuracy. At the same time, the classification performance of the compressed description directly depends on the procedure of its formation. The simultaneous provision of high accuracy and classification performance is achieved by the procedure of stepwise reduction of descriptions for the database etalons based on the evaluation of the informativeness criterion Eq. (7).

It should be noted that in the process of selecting descriptors by informativeness, its actual value changes and needs to be recalculated for further use, since the value of informativeness Eq. (7) is directly affected by the composition of the resulting reduced base.

Another caveat concerns the direct use of information content values in models of the form Eq. (2). The implementation of such models can be effective only if the calculated informativeness values differ significantly within the same etalon or for different etalons. In our experiment, the informativeness values for the selected 50 descriptors ranged from 34...50 for the first etalon and 24...46 for the other four, with approximate average values of 38, 28, 31, 31, 29. As we can see, these values are quite close to each other, which means that the direct implementation of the model Eq. (2) with these values does not make sense, as it will not affect the enhancement of classification indicators.

Therefore, based on the set of 50 selected descriptors for the etalons, we recalculated the values of informativeness for them. They formed the ranges −19...+59, −28...+49, −9...+49, −27...+57, −21...+60 with average values of 36, 31, 33, 30, 38. Thus, the total range of informativeness values for the modified database was −21...+60.

Based on the obtained informativeness indicators, new descriptions of the etalons with 25 descriptors each were formed by selecting the largest values.

The simulation showed that for the description database of 25 descriptors, only 2 false positive objects (assigned to a certain class) were identified from the number of images that are not included in the etalon images. At the same time, all images from the database were classified correctly! The accuracy is 0.95.

Similarly, out of the 50 descriptors selected in the first stage, the 10 most informative descriptors were identified. On the test set, 10 misclassified objects out of 51 were experimentally identified, and the accuracy was 0.81. The decrease in accuracy was more influenced by images not from the database, as the similarity of all images increases significantly for a small sample of features. If we exclude the condition for not exceeding the threshold δ_Δ from the classification rule, the accuracy slightly increases to $pr = 0.93$, which is a fairly high rate.

It should be noted that, according to the principles of data science [17,18], the rather high performance obtained directly depends on the test set of images within and outside the selected database. But in any case, the results of our experiments make it possible to significantly reduce the classification time (for the variant with 25 informative descriptors–by about 20 times!) without significantly reducing the accuracy rate. The most practical way of classification is to make a decision based on half or more of the votes of the generated description of the etalon.

Experiments with the accumulation of the parameter of descriptors' informativeness as a variant of the model Eq. (4) in the decision-making process showed the following.

To simplify the calculations, we divided the information content coefficients listed for the 50 descriptors into 4 intervals of approximately equal width, and assigned them interval weights of 1, 2, 3, 4, so that these weights could be accumulated along with the number of votes, as well as the product of the corresponding weight and the minimum distance obtained by the matching search. The analysis showed that the vast majority of informative etalon descriptors (39...43 out of 50 for different etalons) received interval coefficients of 1 and 2, i.e., have a significant level of information content. This can be explained by the fact that the analyzed data have already passed the selection process according to the criterion of informativeness. As a result of classification by the nearest neighbor modification, we have an example for a transformed image of the 2nd class in terms of accumulated votes [13, 43, 3, 4, 7], values [17, 70, 6, 6, 14], and products of values by distance [938, 2270, 332, 347, 866] with the parameter $\Delta = 0.31$ for votes.

As we can see from this example, the values obtained by the classifier for votes and importance are highly correlated. This can be explained by the fact that the applied informativeness criterion Eq. (7) is defined through the minimum metric, which means that metric correlations dominate in such a classifier. Experimentally, in some cases, for images from the database, the accumulated significance correctly indicates the true class of the object, but for images not from the database, it only confirms the fact that the description is incorrectly assigned to one of the classes by the number of votes.

To summarize the results of the research, we conducted a software modeling of the classifier for the same test data using the Cross-Checking model for double-checking the descriptors' correspondence. For 500 descriptors, the accuracy of 1.0 has not changed, but to ensure it, it is necessary to reduce the threshold δ_h for the number of votes to 220. This is necessary for the correct classification of the 2nd etalon, which is the most similar to the others.

The confidence factor Δ for the images from the database improved to 0.08, and the t_c computation time almost did not increase compared to the approach without double-checking. It is worth noting that the experimental classification time using Cross-Checking varies widely for different test images compared to the first approach, where it is virtually constant.

For a set of 50 descriptors, the maximum accuracy of 1.0 was achieved at the threshold of δ_h of 25 votes for the database, $\Delta \leq 0.1$ for the database, the classification time is about 10 times less than for the full description set. For 25 descriptors, the accuracy of 1.0 was achieved at the threshold of δ_h of 15 votes, $\Delta \leq 0.5$ for the database, the classification time is 2 times less than for the description composition of 50 descriptors. Studying the variant for 25 descriptors using the accumulation of val_k significance confirms the classification using only δ_h.

An important result was obtained for a small number of descriptors of 10 elements, where the significance coefficient may have a greater impact. When choosing the threshold $\delta_h = 8$, as in the first method, the accuracy drops to the level of $pr = 0.84$. However, when choosing the threshold of $\delta_h = 6$, the accuracy improves significantly to $pr = 0.96$ (but with a significant increase in $\Delta \leq 0.85$). An option to improve the result for small descriptions is to make an optimal decision solely with the checking of the parameter δ_h.

At the same time, the classification by the accumulated significance of val_k fully (even with an incorrect decision) confirms the conclusion of the vote count. In this case, it is more important to choose the threshold δ_h than to use the indicator Δ.

When classifying 25 descriptors according to model Eq. (2), where the criterion is the product of significance and distance, the following results were obtained. To achieve the score, the $pr = 1.0$ threshold δ_h should be set to 10, while the computation time should increase slightly. For a description with 10 descriptors, the accuracy when choosing $\delta_h = 4$ is $pr = 0.84$, this is slightly worse than for the traditional classifier by the number of votes. The obtained experimental data are presented in Table 1.

The analysis of the experiments and the content of Table 1 lead to the following preliminary conclusions:

1. Cross-Checking provides a more accurate classification with a better value Δ by reducing the impact of emissions.
2. The accumulated significance is correlated with the number of votes of the classes and can be used to confirm the decision. The classification performance by the accumulated significance almost coincides with the decision by the number of votes.
3. The method of calculating matches with Cross-Checking is more sensitive to the choice of threshold δ_h. The performance of Cross-Checking varies widely depending on the data values.
4. The Δ confidence score deteriorates (increases) with the number of descriptors in the description. The value of Δ with $s = 500$ is ten times better for the method with Cross-Checking.
5. The classification accuracy for the researched methods, especially when describing in 10 descriptors, can be improved by adaptive selection of δ_h and simplifying the classification by eliminating the check for δ_Δ. The use of significance coefficients has less of an impact on improving classification accuracy for images not in the database.
6. Implementation of the model Eq. (2), where the criterion is the product of the metric and the significance, at $s = 25$ provides accurate classification provided that the threshold δ_h of class votes is reduced to $\delta_h = 10$. However, with a small value of $s = 10$, model Eq. (2) does not improve the classifier's performance for the analyzed data.
7. For small reduced volumes of description, it is advisable to apply a simple classification rule based only on the number of votes of the classes.

6 Conclusion

The introduction of significance in the form of the data informativeness criterion into the process of structural classification enhances adaptability with class etalons and ensures informed decision-making. The use of the informativeness parameter opens up new possibilities for managing the process of data analysis in the course of classification. The key to the classification process is the metric relationship between descriptors and etalons.

The significance parameter can be successfully used only in situations where its value has variability across a set of descriptors or etalons. Classification using reduced descriptions of the etalons, compared to the full description, remains effective only if the description is reduced by selecting according to the information content criterion. Directly generating a smaller description significantly worsens the classification accuracy rate.

The use of different ORB and BRISK descriptors in the experiment confirms the universality of the proposed mechanism for reducing description data regardless of the type of detector.

The main result of the research is the establishment of high performance and efficiency of classifiers based on the reduced composition of the description of the etalons. The use of reduction to transform the set of image description descriptors makes it possible to significantly speed up processing without significant loss of classification accuracy. The processing speed increases in proportion to the degree of data reduction and was improved by a factor of 20 in the experiment. It is practically advisable to use 10% of the most informative descriptors, which provides a 10-fold increase in performance while maintaining full accuracy. If speed is a key criterion, then it is permissible to use even 5% of the original number of descriptors, which provides an increase in speed by almost 20 times, but with a slight decrease in accuracy to 0.95.

Acknowledgement: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research (work through the Project Number PSAU/2023/01/25387). The authors acknowledge the support to the undergraduate students Oleksandr Kulyk and Viacheslav Klinov of the Department of Informatics, Kharkiv National University of Radio Electronics, for participating in the experiments and to the management of the Department of Informatics for supporting the research.

Funding Statement: This research was funded by Prince Sattam bin Abdulaziz University (Project Number PSAU/2023/01/25387).

Author Contributions: The authors confirm contribution to the paper as follows: research conception and design: Yousef Ibrahim Daradkeh, Volodymyr Gorokhovatskyi, Iryna Tvoroshenko, Medien Zeghid; methodology: Volodymyr Gorokhovatskyi, Iryna Tvoroshenko; data collection: Volodymyr Gorokhovatskyi, Iryna Tvoroshenko; analysis and interpretation of results: Volodymyr Gorokhovatskyi, Iryna Tvoroshenko; draft manuscript preparation: Iryna Tvoroshenko. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

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

References

[1] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, and M. Zeghid, "Tools for fast metric data search in structural methods for image classification," *IEEE Access*, vol. 10, pp. 124738–124746, Nov. 2022. doi: 10.1109/ACCESS.2022.3225077.

[2] 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. 43, no. 4, pp. 85–96, Mar. 2018. doi: 10.15588/1607-3274-2017-4-10.

[3] H. Kuchuk, A. Podorozhniak, N. Liubchenko, and D. Onischenko, "System of license plate recognition considering large camera shooting angles," *Radioelectron. Comput. Syst.*, vol. 100, no. 4, pp. 82–91, Dec. 2021. doi: 10.32620/reks.2021.4.07.

[4] Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, S. Gadetska, and M. Al-Dhaifallah, "Methods of classification of images on the basis of the values of statistical distributions for the composition of structural description components," *IEEE Access*, vol. 9, pp. 92964–92973, Jun. 2021. doi: 10.1109/ACCESS.2021.3093457.

[5] X. Shen *et al.*, "Learning scale awareness in keypoint extraction and description," *Pattern Recognit.*, vol. 121, no. 2, pp. 108221, Jan. 2022. doi: 10.1016/j.patcog.2021.108221.

[6] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 2548–2555.

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

[8] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Distance-based image classification: Generalizing to new classes at near-zero cost," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2624–2637, Nov. 2013. doi: 10.1109/TPAMI.2013.83.

[9] Q. Wu, Y. Li, Y. Lin, and R. Zhou, "Weighted sparse image classification based on low rank representation," *Comput. Mater. Contin.*, vol. 56, no. 1, pp. 91–105, Jul. 2018. doi: 10.3970/cmc.2018.02771.

[10] A. Nasirahmadi and S. -H. M. Ashtiani, "Bag-of-feature model for sweet and bitter almond classification," *Biosyst. Eng.*, vol. 156, no. 27, pp. 51–60, Apr. 2017. doi: 10.1016/j.biosystemseng.2017.01.008.

[11] 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, no. 10, pp. 126938–126949, Nov. 2023. doi: 10.1109/ACCESS.2023.3332291.

[12] Z. Wang, X. Zhao, and Y. Tao, "Integrated algorithm based on bidirectional characteristics and feature selection for fire image classification," *Electronics*, vol. 12, no. 22, pp. 4566, Nov. 2023. doi: 10.3390/electronics12224566.

[13] P. Flach, "Features," in *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, New York, NY, USA: Cambridge University Press, 2012, pp. 298–329.

[14] V. O. Gorokhovatskyi, I. S. Tvoroshenko, and O. O. 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.

[15] V. Gorokhovatskyi, O. Peredrii, I. Tvoroshenko, and T. Markov, "Distance matrix for a set of structural description components as a tool for image classifier creating," *Adv Inf. Syst.*, vol. 7, no. 1, pp. 5–13, Mar. 2023. doi: 10.20998/2522-9052.2023.1.01.

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

[17] J. Leskovec, A. Rajaraman, and J. D. Ullman, "Finding similar items," in *Mining of Massive Datasets*, New York, NY, USA: Cambridge University Press, 2020, pp. 73–128.

[18] U. Stańczyk, "Feature evaluation by filter, wrapper, and embedded approaches," in *Feature Selection for Data and Pattern Recognition*, Berlin-Heidelberg, Germany: Springer, 2014, vol. 584, pp. 29–44.

[19] M. S. Biagio, L. Bazzani, M. Cristani, and V. Murino, "Weighted bag of visual words for object recognition," in *2014 IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2734–2738.

[20] 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, Oct. 2020. doi: 10.1109/ACCESS.2020.3028740.

[21] V. Gorokhovatskyi, I. Tvoroshenko, O. Kobylin, and N. Vlasenko, "Search for visual objects by request in the form of a cluster representation for the structural image description," *Adv. Electr. Electron. Eng.*, vol. 21, no. 1, pp. 19–27, May 2023. doi: 10.15598/aeee.v21i1.4661.

[22] J. Leskovec, A. Rajaraman, and J. D. Ullman, "Clustering," in *Mining of Massive Datasets*, New York, NY, USA: Cambridge University Press, 2020, pp. 241–276.

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

[24] Y. Liu, Y. Pan, and J. Yin, "Deep hashing for multi-label image retrieval with similarity matrix optimization of hash centers and anchor constraint of center pairs," in *Communications in Computer and Information Science*, Singapore, Springer, 2023, vol. 1967, pp. 28–47.

[25] S. Mashtalir and V. Mashtalir, "Spatio-temporal video segmentation," in *Studies in Computational Intelligence*, Cham, Switzerland, Springer, 2019, vol. 876, pp. 161–210.

[26] H. Jegou, M. Douze, C. Schmid, and P. Perez, "Aggregating local descriptors in to a compact image representation," in *2020 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 3304–3311.

[27] Y. I. Daradkeh and I. Tvoroshenko, "Application of an improved formal model of the hybrid development of ontologies in complex information systems," *Appl. Sci.*, vol. 10, no. 19, pp. 6777, Sep. 2020. doi: 10.3390/app10196777.

[28] V. Gorokhovatskyi, S. Gadetska, and R. Ponomarenko, "Recognition of visual objects based on statistical distributions for blocks of structural description of image," in *Proc. XV Int. Sci. Conf. Intell. Syst. Decis. Making Problems Comput. Intell, (ISDMCI'2019)*, Ukraine, 2019, vol. 1020, pp. 501–512.

[29] Y. Wang and P. Chen, "A lightweight multi-scale quadratic separation convolution module for CNN image-classification tasks," *Electronics*, vol. 12, no. 23, pp. 4839, Nov. 2023. doi: 10.3390/electronics12234839.

[30] D. P. 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, Mar. 2023. doi: 10.1109/ACCESS.2021.3137638.

[31] Z. Hu, Y. V. Bodyanskiy, and O. K. Tyshchenko, "A deep cascade neural network based on extended neo-fuzzy neurons and its adaptive learning algorithm," in *2017 IEEE First Ukraine Conf. Electr. Comput. Eng. (UKRCON)*, Kyiv, Ukraine, Jun. 2017, pp. 801–805.

[32] C. B. Murthy, M. F. Hashmi, G. Muhammad, and S. A. AlQahtani, "YOLOv2PD: An efficient pedestrian detection algorithm using improved YOLOv2 model," *Comput. Mater. Contin.*, vol. 69, no. 3, pp. 3015–3031, Aug. 2021. doi: 10.32604/cmc.2021.018781.

[33] O. Yakovleva and K. Nikolaieva, "Research of descriptor based image normalization and comparative analysis of SURF, SIFT, BRISK, ORB, KAZE, AKAZE descriptors," *Adv. Inf. Syst.*, vol. 4, no. 4, pp. 89–101, Dec. 2020. doi: 10.20998/2522-9052.2020.4.13.

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

[35] M. A. Ahmad, V. Gorokhovatskyi, I. Tvoroshenko, N. Vlasenko, and S. K. Mustafa, "The research of image classification methods based on the introducing cluster representation parameters for the structural description," *Int. J. Eng. Trends Technol.*, vol. 69, no. 10, pp. 186–192, Oct. 2021. doi: 10.14445/22315381/IJETT-V69I10P223.

[36] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021. doi: 10.1109/TPAMI.2020.2970919.

[37] P. Wei, J. Ball, and D. Anderson, "Fusion of an ensemble of augmented image detectors for robust object detection," *Sensors*, vol. 18, no. 3, pp. 894, Mar. 2018. doi: 10.3390/s18030894.

[38] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *VISAPP Int. Conf. Comput. Vis. Theory Appl.*, Lisboa, Portugal, 2009, pp. 331–340.

[39] R. Szeliski, "Recognition," in *Computer Vision: Algorithms and Applications*, London, UK: Springer-Verlag, 2010, pp. 655–718.

[40] A. Alzahem, W. Boulila, A. Koubaa, Z. Khan, and I. Alturki, "Improving satellite image classification accuracy using GAN-based data augmentation and vision transformers," *Earth. Sci. Inform.*, vol. 16, no. 4, pp. 4169–4186, Nov. 2023. doi: 10.1007/s12145-023-01153-x.

[41] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analisis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1/2, pp. 23–69, Oct. 2003. doi: 10.1023/A:1025667309714.

[42] G. Bradski and A. Kaehler, "OpenCV open source computer vision," 2024. Accessed: Jan. 12, 2024. [Online]. Available: https://docs.opencv.org/master/index.html

[43] S. Walt *et al.*, "ORB feature detector and binary descriptor," Dec. 24, 2023. Accessed: Jan. 16, 2024. [Online]. Available: https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_orb.html

[44] E. Karami, S. Prasad, and M. Shehata, "Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images," arXiv preprint arXiv:1710.02726, 2017.

[45] C. Battet and E. Durgoni, "Google arts & culture," 2023. Accessed: Jan. 26, 2024. [Online]. Available: https://artsandculture.google.com/partner/national-museum-of-the-history-of-ukraine