V. Gorokhovatskyi
Doctor of technical sciences, professor
I. Tvoroshenko
Candidate of technical sciences, associate professor
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

AN EFFECTIVE METHOD FOR TRANSFORMING AN IMAGE DESCRIPTION INTO A COMPACT VECTOR FOR CLASSIFICATION

Abstract. The work develops method to solve a fundamental problem in computer vision: image recognition of visual objects. Based on the implementation of the distance matrix model, it was possible to form effective integrated features in the form of one-dimensional data distributions and vectors for the sum of the matrix columns, which reduced computational costs without losing the effectiveness of classification on the training data sample. The efficiency of image classification was experimentally evaluated using software modeling.
Keywords: computer vision, descriptor, distance matrix, keypoint, orthogonal decomposition, processing speed, relevance of descriptions.

In structural methods of classification, the image of an object is represented by a finite set of vectors that are descriptors of keypoints of the image [1]. Methods using descriptors of keypoints are physically based on information and features of the image itself. The value of the relevance measure for a pair of image descriptions is estimated as the proximity or distance between two sets of vectors of approximately the same power [1]. The basis for determining such a measure is the distance matrix, which contains the value of the metric between all pairs of elements of two sets according to the “one-to-one” rule.

The tasks of the research consist in implementing orthogonal decomposition of the etalon and input images, applicating the distance matrix apparatus to obtain integral characteristics of the description, constructing a metric classifier in the transformed space of features, studying the effectiveness of the developed classifiers’ modifications (performance, speed, immunity) by means of simulation modeling.

One of the directions to develop structural methods is the conversion of the existing features’ system (a set of descriptors) by creating a modified space with the necessary classification properties. Such structural description conversions as a hierarchical representation [2], cluster representation [3], division of descriptors into groups by distance to the medoid [4], spatial analysis of the system of non-intersecting fragments [5] have been proposed and successfully implemented in the classification problem. Such conversions not only speed up the operation of the classifier, but also ensure a sufficient level of accuracy.

We conclude that the conceptual decomposition of the component-descriptors’ set and the formalism of the distance matrix for the composition of class descriptions can be the sources to build a new effective system of classification features. It is proposed to study the possibility and effectiveness of their application in structural classification models.

In the space of structural features, the image description $Z=\{z_i\}_{i=1}^{s}$ is a dimensional vector space with s vectors Z_i, $s = card\,Z$, $z_i \in B^n$, $z_i = \{z_{i,j}\}_{j=1}^{n}$, B^n – vector space of dimension n with binary components {0,1} [1], [5]. We consider the descriptor of the image keypoints as a point of n-dimensional discrete signal.

We will interpret the conversion of the description Z as a mapping $B^n \to R$, where R is the newly created space of classification features.

In any finite vector space there is at least one orthogonal system of vectors $W = \{\{w_{j,b}\}_{j=1}^{n}\}_{b=1}^{n}$, so that any vector $z \in B^n$, $z = \{z_b\}_{b=1}^{n}$ can be represented in it by a tuple $\alpha = \{\alpha_j\}_{j=1}^{n}$, where

$$\alpha_j = (z \cdot w_j) = \sum_{b=1}^{n} z_b w_{j,b}\,. \quad (1)$$

Coefficient (1) is defined as the scalar product of a vector z on the vector w_j orthogonal system [3]. Parameters α_j are the vector's coordinators in base W. For each $z_i \in Z$ we have the value of the expansion vector α_i. It follows from the introduction of the vector space model to classification that any descriptor $z_i \in Z$ of an image illustrated can be inversely presented in a single way as a linear combination of basis vectors, and the selected basis constitutes exactly n vectors, i.e.

$$z_i = (\alpha_i \cdot w_j) = \sum_{b=1}^{n} \alpha_{i,b} w_{j,b}\,. \quad (2)$$

Note that in n-dimensional space of vectors theoretically there is a set of different orthogonal bases [3]. Considering the fact that the Walsh functions formally take the value {1, -1} and do not directly enter the space B^n, reconstruction (1) actually reflects the conversion $B^n \to C^n$ into space C^n vectors with integer components. Just at the same time space $B^n \subset C^n$ is subspace for space C^n, therefore, it can be assumed that the conversions (1), (2) are carried out in space C^n. The normalization coefficient for the Walsh functions to the orthonormal system, which provides the inverse of the conversions (1), (2), is a constant for a given n, so it can be ignored when calculating the features. As a result of applying component-wise conversion $z_i \to \alpha_i$ description Z will look as $\alpha(Z) = \{\alpha_i\}_{i=1}^{s}$, i.e. a set of keypoints descriptors $\{Z_i\}_{i=1}^{s}$, $z_i \in B^n$ is transformed into a set of $\{\alpha_i\}_{i=1}^{s}$, $\alpha_i \in C^n$, of the same size and number. The question arises about the effectiveness of such a conversion in the implementation of classification.

Conversions (1), (2) can be written in vector form using Hadamard matrix A. The complete set of Walsh functions of dimension n forms an orthogonal Hadamard matrix A in size of $n \times n$, which consists of vectors of n Walsh functions vectors $W_1,...,W_n$. We apply the decomposition of the description in the basis of a family of rectangular basis functions – discrete Walsh functions, which are vectors of integers (value +1, -1) of finite dimension (power of two) [3]. We regard description Z as rectangular matrix $Z = \{z_{i,j}\}$, $i = \overline{1,s}$, $j = \overline{1,n}$ of size $s \times n$, whose rows contain s descriptors. Conversion Z can be given as multiplication of rectangular and square matrices

$$U = Z \times A. \tag{3}$$

As a result of (3), we get a rectangular matrix $U = \{u_{i,j}\}$, $i = \overline{1,s}$, $j = \overline{1,n}$ in size of $s \times n$, the rows of which contain expansion vectors α for Z description descriptors. In our research the set of Walsh functions is considered and applied in full, we will not analyze the variety of ways to arrange Walsh functions systems [3].

Traditionally, classification consists in establishing the degree of relevance $\Theta(Z, E_k)$ between the description $\{Z_i\}_{i=1}^{s}$ of the analyzed image and the components $E_k = \{e_i(k)\}_{i=1}^{s}$ of the database $E = \{E_k\}_{k=1}^{N}$ etalon descriptions [1]. The class m of the object is determined according to the extremum of the function $\Theta(Z, E_k)$ as

$$m = \arg \underset{k=1,\dots,N}{extr} \Theta(Z, E_k). \tag{4}$$

After the conversion, the classification process will be based on the definition of the modified relevance measure $\Theta^{\alpha}(\alpha(Z), \alpha(E_k))$ for transformed descriptions $\alpha(Z)$ and $\alpha(E_k)$ in the new feature space. Each of the transformed descriptions is a set of vectors with components from C^n. Seeing the available possibility of data reduction in the newly created space, we will also have in mind some reduced space C^{α}, obtained from C^n by reducing the number of applied Walsh functions to $q << n$. Paying attention to the limited range of values of the input signal and the controlled conversion in the space of integers, when building a classifier, it is possible to accurately estimate the range of values of functions Θ and Θ^{α}, and use this knowledge for classification. One of the options either Θ or Θ^{α} is metric for sets of vectors. This can be the Tanimoto distance for sets A, B, which contains the ratio of the number of elements of a symmetric difference and the union of sets [3]

$$T(A, B) = \frac{card(A \Delta B)}{card(A \cup B)}. \tag{5}$$

Measure (5) reflects the quantitative characteristics of equivalent and different elements of the compared sets. When applying the metric (5), its limited value is important, for which the elements of the vector space are considered equivalent [3]. This problem is one of the key ones in multidimensional data spaces. The choice of the equivalence threshold significantly affects the classification result.

We offer two ways to solve it. One of them determines limit value

$$\rho_{\lim}(a,b) = \alpha \rho_{\max}(a,b), \tag{6}$$

as a percentage α of the theoretically determined maximum of the metric. For example, for the Hamming metric, which for vectors in size of 256 bits varies in the interval [0, 256], with a percentage of $\alpha = 25\%$ it is possible to determine $\rho_{\lim}(a,b) = 0.25 * 256 = 64$. The method of obtaining $\rho_{\lim}(a,b)$ is more practical according to the result of the analysis of applied experimental data, where $\rho_{\max}$ is calculated for a set of etalons. Calculation time of $\rho_{\max}$ does not affect the costs for classification and is carried out at the previous stage of data analysis.

We will evaluate the performance of the classification method by the accuracy indicator pr, which is calculated by the ratio of the number of correctly classified objects r_p to their total number r, that was used in the experiment [3]

$$pr = r_p / r. \quad (7)$$

Immunity is also an important indicator of the effectiveness of image recognition methods. It is characterized by the value of the accuracy of classification in the conditions of interference [6]. If the effect of additive interference on the image $B(x, y)$ is described by the model $B_\xi(x, y) = B(x, y) + \xi(x, y)$, and trouble $\xi(x, y)$ is characterized by the root mean square deviation σ with zero mathematical expectation, then the signal-to-noise ratio is described as $\mu = B_m / \sigma$, where B_m – amplitude characteristic (e.g., the average brightness value). An indicator of immunity is dependence $pr(\mu)$ for the developed method.

The novelty of the research is the improvement of the structural method of image classification based on the implementation of description conversion through orthogonal decomposition of data and the construction of feature models based on distance matrices in the descriptor space, as well as methods of compressing descriptions in newly created spaces, which reduces the computational cost of classification.

The results of this research were obtained under the international research project INITIATE under the grant No. 101136775-HORIZON-WIDERA-2023-ACCESS-03.

References:

1. 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.” Computers, Materials & Continua 80.2 (2024): 3085-3106.

2. Y. I. Daradkeh, V. Gorokhovatskyi, I. Tvoroshenko, and M. Al-Dhaifallah. “Classification of images based on a system of hierarchical features.” Computers, Materials & Continua 72.1 (2022): 1785-1797.

3. V. Gorokhovatskyi, I. Tvoroshenko, and O. Yakovleva. “Transforming image descriptions as a set of descriptors to construct classification features.” Indonesian Journal of Electrical Engineering and Computer Science 33.1 (2024): 113-125.

4. I. S. Tvoroshenko, and V. O. Gorokhovatsky. “Intelligent classification of biophysical system states using fuzzy interval logic.” Telecommunications and Radio Engineering 78.14 (2019): 1303-1315.

5. 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 12 (2024): 73376-73385.

6. 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 11 (2023): 126938-126949.