IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Quantization of the Space of Structural Image Features as a Way to Increase Recognition Performance

Volodymyr Gorokhovatskyi
Informatics department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
gorohovatsky.vl@gmail.com

Oleksii Gorokhovatskyi
Informatics and Computer Technologies department
Simon Kuznets Kharkiv National University of Economics
Kharkiv, Ukraine
oleksii.gorokhovatskyi@gmail.com

Putyatin Yevgenyi
Informatics department
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine

Peredrii Olena
Informatics and Computer Technologies department
Simon Kuznets Kharkiv National University of Economics
Kharkiv, Ukraine

***Abstract* — A modification of the structural image recognition method in computer vision systems is proposed. In order to improve the performance of recognition, quantization (clustering) is applied in the space of characteristic features that form the pattern of the object. Due to the transformation of structural objects descriptions from a set representation to a vector form, the amount of computation is reduced tens of times. The results of experiments that confirmed the effectiveness and increase of decision-making process are shown.**

***Keywords* — computer vision; structural recognition methods; set of characteristic features; descriptor; quantization; competitive learning; recognition performance; noise immunity**

I. Introduction

Structural methods for image recognition, where visual objects are represented by sets of characteristic features (points of interest), are widely used due to high efficiency associated with the practical ability to perform the investigation and recognition in complex conditions connected with the presence of a partial representation of the patterns [1-3]. In this case, the user of computer vision system has the ability to determine independently which specific part of the available description is acceptable for decision-making about the class of an object. Successful implementations of structural methods in problems of face recognition, animals recognition, as well as a number of iconic types images are known: coats of arms, paintings, logos, brands, etc. [4, 5].

The main factors that determine the effectiveness of computer recognition are the features detection method, the space of images within decision is made of, as well as the level of influence by external obstacles (noise, distortions). Usage of information about etalons during recognition process determines the quality of learninh (tuning) of the system for specific initial data. Embedding of the learning step allows not only to improve recognition performance by adaptation of parameters but also ensures the universality and the ability of its functioning for arbitrary image databases [4].

An effective way to increase the effectiveness of the structural recognition methods in terms of speed and, in fact, without reducing the probabilistic characteristics, is the use of the vector quantization (granulation) in the space of structural features [4]. Vector quantization allows to approximate the space of key point descriptors by splitting them into subsets of equivalent elements. Due to the quantized representation, with the creation of set-vector mapping, the features space is transformed, as a result, the calculation of relevance of objects descriptions can be interpreted as distance or vector similarity [3]. Vector quantization without a teacher (self-learning) automatically classifies input elements, which makes it possible to apply it for a number of applications not related to image analysis, for example, in the field of digital telecommunications.

Applying quantization in the space of characteristic features (CF) of images allows to obtain subsets of close elements represented by a set of centers, and out of them, like of bricks, we form the image of an arbitrary visual object. Such models can be classified as structural-statistical. Formation of object's CF, for example, by SURF, ORB detectors, allows reaching invariance to permissible geometric transformations [2, 7].

The procedures for quantization and learning as its part in the organization of recognition process are often interpreted by researchers as a whole. However, vector quantization itself is, in our opinion, more objective, and learning, having a specific nature of the object of research, is more aimed to meaningful management of the adaptation process.

Goal of the work is to investigate and evaluate the effectiveness of quantization in structural recognition of images based on image descriptions in the form of sets of CF.

The objectives of investigation are the analysis of quantization qualities and learning process for the list of vectors-descriptors of images, as well as the evaluation of the effectiveness of learning and recognition for real image databases.

978-1-5386-2874-4/18/$31.00 ©2018 IEEE

II. Quantizing of the Descriptor Space

Accordingly to VQ (Vector Quantization), we perform a discrete approximation of continuous (in the general case) input data from the set of vectors $W=\{\mathrm{x} \mid \mathrm{x}\in \mathrm{R}^n\}$, $\mathrm{W}\subseteq R^n$ using predefined set of k encoding vectors $\mathrm{M}=\{\mathrm{m}_i\}_{i=1}^k$, $m_i \in R^n$, $i=1,2,...,k$ [6,8]. Concurrent approximation of vector $x\in W$ means a search for a number υ of vector $m_\upsilon \in M$ closest to it (in terms of Euclid $\rho(\mathrm{x},\mathrm{m}_i)$ metrics usually) in the space of encoding vectors:

$$\upsilon = \arg \min_{i=1,...,k} \rho(\mathrm{x},\mathrm{m}_i)\,. \tag{1}$$

Model (1) in known as concurrent Kohonen learning [6]. Idea of a quantization is the formation of M on the basis of training set W accordingly to optimum of some criteria.

Normalization of W is required to ensure stable stationary learning process and equivalent influence of input vectors to result: $W=\{\mathrm{x}^* \mid \mathrm{x}^*=\mathrm{x}/||\mathrm{x}||,\ ||\mathrm{x}||=\sqrt{\sum_i \mathrm{x}_i^2}\}$, assuming: $||\mathrm{x}^*||=1$.

In the space of CF formed by SURF, ORB feature detectors, this condition is fulfilled, so no additional signal normalization is required.

Quantization in general form can be formulated as a global optimization problem for some functional, that reflects the quality of the cluster system, while the total distances between the elements within the clusters are minimized, and the distances between the cluster centers are maximized. Taking into account, that quantization plays an auxiliary role here, the key criterion is the probability of correct recognition.

Batch processing during quantization is implemented in the form of a computational scheme, which is used in a situation where the set W is available as a whole at the beginning of learning, and all $x\in W$ are considered to be equivalent. A common version of batch processing – k-means (C-means) [6] is applicable for arbitrary metrics when comparing elements. The k-means algorithm arranges cluster centers (centroids) so that the average values for the lists of elements within the clusters differ as much as possible.

Stages of calculations applied to sets of CF:

1) random k vectors are assigned as initial centers $M=\{m_i\}_{i=1}^k$, $m_i \in W$;

2) for each $i=\overline{1,k}$ with learning according to (1) list $W_i \subseteq W$ is formed. This list contains elements, which have m_i as nearest encoding vector, i.e. form subsets $W_i=\{x\in W \mid arg\ \min_\upsilon \rho(x,m_\upsilon)=i\}$; list $\{W_i\}$ in this case creates partition W : $W=\cup W_i$, $W_i \cap W_j = \varnothing$;

3) average value according to W_i is calculated as next m_i value: $m_i=\sum_{\upsilon=1}^{s(i)} x_\upsilon / s(i)$, where $s(i)=card W_i$ – amount of elements in W_i;

4) steps 2 and 3 are repeated while list in unstable in terms of some criterion.

This algorithm is especially effective if the initial values of the vectors m_i are previously somehow coordinated with the training set. The algorithm does not contain the learning speed parameter, which is not required to be controlled during processing, and there are no problems with convergence. Algorithm is stopped if changes of the centers $\{m_i\}$ become insignificant between iterations, which is evaluated by criteria:

$$\Delta(M[h+1],M[h]) \le \varepsilon_M\,,$$

where Δ is a measure of difference between two lists of centroids, $M[h]=\{m_i(h)\}_{i=1}^k$ are the values of list on step h of an iteration; ε_M – threshold value a priori.

As an example of Δ criteria sum of distances between lists may be used $\Delta=\sum_{i=1}^k \rho(m_i(h+1),m_i(h))$.

In order to make calculations more effective from performance point of view Δ value during each iteration in calculated only for centroids, that changes their value.

Experimental modeling of k-means scheme described above for sets of CF shows, that it is enough to do only a few iterations [4, 9].

K-means algorithm approximates the distribution function of the set of input samples by the criterion of minimum error, that is usually the sum of squares of deviations from cluster centers [8]:

$$E=\sum_{i=1}^k \sum_{\upsilon=1}^{s(\mathrm{i})} \rho^2(\mathrm{x}_\upsilon, m_i)\,, \tag{2}$$

where $s(\mathrm{i})$ is a power of i cluster.

Distance $\rho(\mathrm{x},m_i)=\min_{i=1,...,k}\rho(\mathrm{x},\mathrm{m}_i)$ between element and its cluster center is quantization error. Iterative k-means algorithm converges to local minimum of error E. Minimization of (2) promotes quantization process to fit training data in a best way.

Numerous modifications of k-means method are known, e.g., k-median method, where in order to remove anomaly values median value $\mathrm{m}_i = med\ \{x_\upsilon\}_{\upsilon=1}^{s(\mathrm{i})}$ is selected instead of average in each cluster. The median is defined as an element of a set whose total distance to the remaining elements is minimal [6]. If it is necessary to analyze overlapping clusters, where the values of the membership function for each of the clusters are calculated, the methods of Fuzzy Classifier Means [8] are applicable.

In case of online-learning, when CF $\mathrm{x[t]}\in \mathrm{W}$ comes into processing one by one, center m_υ of cluster, that is the winner in (1) is corrected in a way:

$$\mathrm{m}_\upsilon[\mathrm{t}+1]=\mathrm{m}_\upsilon[\mathrm{t}]+\alpha[\mathrm{t}](\mathrm{x}[\mathrm{t}]-\mathrm{m}_\upsilon[\mathrm{t}])\,, \tag{3}$$

on learning step $t=1,2,...,s$, where $s = card\ W$ is a size of training set (total quantity of CF for the whole dataset), $\alpha[\mathrm{t}]$ is set up by researcher and specifies learning speed assuming $\alpha[\mathrm{t}]\to 0$ and $\mathrm{t}\to s$.

A huge variety of learning strategies and methods has been developed (3), including modeling of dynamics of network topology [8].

III. PROPERTIES OF STRUCTURAL IMAGE DESCRIPTIONS LEARNING

Characteristic features is a vector size of n (usually 32 or 64), that is calculated with usage of some detector to image brightness function. ORB and SURF [2, 7] detectors are the most widespread. Structural description of image is a finite set $O \subset R_1^n$, $R_1^n = \{z \mid z \in \mathrm{R}^n, \| z \| = 1\}$, where $R_1^n \subset R^n$ – is a subset of n-dimensional real vectors that have norm $\| z \| = 1$ [2]. Implementation of the normalization condition allows usage of CF descriptors in the learning procedures directly.

During preprocessing stage set $\mathrm{Z}=\{\mathrm{Z}^j\}_{j=1}^J$ of dataset image features, that includes all patterns (Z^j – is an etalon, J – is the amount of classes), is split on finite amount of k clusters $\mathrm{M}=\{\mathrm{M}_i\}_{i=1}^k$, in a way that $\mathrm{M}_i \cap \mathrm{M}_d = \varnothing$, $M = Z$, clusters are defined with centers $\mathrm{m}=\{\mathrm{m}_i\}_{i=1}^k$. Clustering maps set of CF of the whole dataset into itself $Z \to Z$, each CF belongs to just one cluster. As a result of clustering situation $\mathrm{m}_i \notin W$ is possible. Sets Z and Z^j are multisets, where close CF we consider as equivalent. After clustering is completed, we perform the "screening" of each etalon pattern, as a result, the description Z^j of etalon takes the form:

$$\mathrm{H}[\mathrm{Z}^j] = (\mathrm{h}_1, \mathrm{h}_2, ..., \mathrm{h}_i, ..., \mathrm{h}_k)^j , \quad (4)$$

where $h_i = card\{\mathrm{z} \mid \mathrm{z} \in \mathrm{Z}^j \,\&\, \mathrm{z} \in M_i\}$, $h_i \in C$ – amount of elements of etalon Z^j, that belong to M_i cluster, C – is a set of integers.

An interesting case is a quantization with the number of centers equal to the number of classes to recognize ($k = J$). At the learning stage, for each etalon, the "centers" of the attribute descriptions of the class are formed, and we can directly perform the recognition of a visual object, without going to clusters, in real time, because condition $\mathrm{Z}^j = M_i$ is already established. Here, the quantization apparatus works as a tool for extracting the most significant distinguishing features of etalons.

The method of recognition of visual object represented by the description $O = \{o_l\}$ for $k = J$ can be reduced to counting of the number of voices of elements in accordance with the competitive rule of optimal closeness in the cluster system

$$o_l \to M_i \mid \arg \min_{d=1,...,J} \rho(o_l, m_d) = \mathrm{i} . \quad (5)$$

As a result of calculation (5) $\forall o_l \in O$ we obtain an integer vector of class voices $q = (\mathrm{q}_1, \mathrm{q}_2, ..., \mathrm{q}_J)$ whose maximum component determines the class d for description O:

$$d = \arg \max_j \mathrm{q}_j \quad (6)$$

At the same time, the usage of two methods proposed earlier [3,4] is possible, the first one is based on construction of an integral vector representation for the object O, and the second one is based on the summation of the vectors of the specific weights of elements classified according to rule (5) to the nearest of the clusters.

The essence of learning for the system of structural recognition is reduced to such problems as the construction of a cluster system for a set of attributes of the image database, rational from recognition efficiency point of view, and the estimation of optimal threshold parameters for the classification of objects [10, 13-15].

Possibility of applying the batch mode (the entire W set is defined) may be considered as characteristic of learning in the CF space, as well as the potential of using the learning with the teacher, since the belonging of individual CF etalons from the image database is known, and each of etalons forms its own class. Other characteristics of recognition include the introduction of filtering individual CF identified as noise, as well as the requirement to equalize the number of features in the descriptions of etalons. Random selection or special procedures to filter "significant" features may be used for such equalization.

IV. PERFORMANCE ANALYSIS

The preliminary calculation of the number of computational operations shows that the gain in recognition speed when applying a cluster representation of the form (4) in comparison with the traditional voting of the entire set of CF of object and etalon directly proportional to the number of etalons and inversely proportional to the number of clusters. Specifically, for SURF characteristics, this advantage can be estimated by the value $\alpha \approx 64s / (64k + k^3)$ that for practical values $k = 8$ (the number of clusters) and $s = 300$ (the number of CF in etalon) is estimated as $\alpha \approx 19$. As you can see, the performance boost is dozens of times. The experiments confirm these calculations.

V. EXPERIMENTS

The confirmation of effectiveness of the proposed method is its effective work in applications [4, 9]. Software package was developed in the form of a web-service that implements the formation of CF-descriptors and image processing with the OpenCV library under the control of the Python server. The client application allows to provide input, recognition and clustering using the k-means method. Software simulation of the batch learning of the Kohonen network was performed for the situation when the number of etalons and clusters is the same, which makes it possible to implement recognition by direct assignment of CF to the etalon on the basis of models (5), (6). Quantization at the stage of preliminary processing was done by cluster transformation of the general content of all structural descriptions of etalons in dataset.

Table 1 shows an example of the distribution of CF SURF for each of the 4 etalons (images of dolphins) in the cluster system. Fig. 1 shows one of the images and the coordinates of its CF, generated by the SURF method.

The key factors for recognition efficiency are the distribution of descriptors over the cluster system, as well as the distribution of cluster elements within each etalon (Table 1). This composition depends on the method of clustering and learning technology. On the one hand, uniform distribution across clusters provides an equivalent representation of the characteristics during decision making. On the other hand, the diversity of the distribution structure due to the predominance of some clusters over others contributes to the improvement of the quality of distinguishing objects from their descriptions.

TABLE I. SPLIT BY CLUSTERS

Etalons	Clusters			
	M_1	M_2	M_3	M_4
Z^1	35	9	17	13
Z^2	69	31	24	53
Z^3	67	14	68	6
Z^4	26	5	15	8

Fig. 1. Dolphin image and set of features (SURF)

Experiments showed that the modified method based on the cluster description in terms of recognition efficiency is almost inferior to the traditional approach with the calculation of the distance between sets. An exception is the case of an insignificant number of clusters (2-3).

Noise resistance of the method of structural recognition based on the vector transformation of descriptions is not inferior, and in some situations, for example, with additive noise, even higher than the traditional approach with voting. Proposed approach has sufficiently high noise immunity: with distortion of up to 30% of the total number of CF from the analyzed descriptions, the method provides an almost error-free recognition with a probability higher than 0.98 within the studied dataset.

Looking at comparison between the effectiveness of the use of SURF and ORB detectors, we note that ORB releases about twice as many CF, however, their dimension is smaller (32 versus 64 for SURF). A lot of SURF detectors "cover" the image of the object in more detail, displaying the features of its shape, while the ORB signs are "grouped" and often focus on the boundaries of the object, which is ineffective.

Our experiments confirmed that the implementation of the ORB detector using OpenCV and C ++ is approximately 10-20 times faster than SURF and takes up about 0.006 seconds for one image. The choice of the detector is entirely determined by the type of images analyzed and the requirements for the applications.

VI. CONCLUSIONS

Quantization of the features space reduces the dimension of the recognition problem and provides adaptation to the image dataset. Proposed approach to the construction of methods for structural image recognition based on the quantization of structural descriptions and the transition to the space of descriptor vectors has the prospect of being used due to higher speed and maintaining a sufficient level of correct recognition and noise immunity. The further development of the approach can be the construction of a cluster system and the implementation of learning procedures within each of etalons, which should ensure a more careful accounting of the properties and improve the discernibility of the processed visual objects.

REFERENCES

[1] V. Gorokhovatsky, “Structural analysis and intellectual data processing in computer vision,” SMIT, Kharkov, 2014. (in Russian).

[2] H. Bay, T. Tuytelaars and L.Van Gool, “Surf: Speeded up robust features,” in: 9th European Conference on Computer Vision, Graz, Austria, pp. 404– 417, May 7-13, 2006.

[3] V. Gorokhovatsky, “Efficient Estimation of Visual Object Relevance during Recognition through their Vector Descriptions,” in: Telecommunications and Radio Engineering, Vol. 75, No 14, pp. 1271–1283, 2016.

[4] V. Gorokhovatsky, A. Gorokhovatsky and A.Berestovsky, “Intellectual Data Processing and Self-Organization of Structural Features at Recognition of Visual Objects,” in: Telecommunications and Radio Engineering, Vol. 75, No 2, pp. 155–168, 2016.

[5] L. Shapiro and G. Stockman, “Computer vision,” Prentice Hall., 2001.

[6] T. Kohonen, “Self-Organizing Maps,” Springer Science & Business Media, 2001.

[7] E. Karami, S. Prasad and M. Shehata, “Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images,” [Online]. Available: https://www.researchgate.net/publication/292157133_[July 30, 2017].

[8] S. Osovski, “Sieci neuronovane do przetwarzania informacji,” in: Oficyna Wydawnicza Politechniki Warszawskiej, 2000.

[9] V.Gorokhovatsky, Y. Putyatin and V. Stolyarov, “Research of Effectivenss of Structural Image Classification Methods using Cluster Data Model,” in: Radio Electronics, Computer Science, Control, vol. 3 (42), pp. 78–85, 2017.

[10] O. Gorokhovatskyi, “Neocognitron As a Tool for Optical Marks Recognition,” in: First IEEE International Conference on DataStream Mining & Processing (DSMP), Lviv, Ukraine, pp. 169 – 172, 23-27August 2016.

[11] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, “ORB: an efficient alternative to SIFT or SURF,” in: IEEE International Conference on Computer Vision (ICCV), pp. 2564 – 2571, 2011.

[12] OpenCV Library [Online]. Available: http://opencv.org [November 15, 2017]

[13] M. Sonka, V. Hlavac and R. Boyle, “Image Processing, Analysis and Machine Vision,” Thomson, 2008.

[14] R. Szeliski, “Computer Vision: Algorithms and Applications,” Springer, 2010.

[15] Y. Amit, “2D Object Detection and Recognition: models, algorithms and networks,” The MIT Press, 2002.