MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SUMY STATE UNIVERSITY
UKRAINIAN FEDERATION OF INFORMATICS

PROCEEDINGS
OF THE V INTERNATIONAL SCIENTIFIC CONFERENCE

ADVANCED INFORMATION SYSTEMS AND TECHNOLOGIES

AIST-2017
(Sumy, May 17–19, 2017)

SUMY
SUMY STATE UNIVERSITY
2017

The Vth International Conference «Advanced Information Systems and Technologies, AIST 2017»
17-19 May 2017, Sumy, Ukraine

Feature learning for information-extreme classifier

A. Korobov, A. Moskalenko, V. Moskalenko
Sumy State University, v.moskalenko@cs.sumdu.edu.ua

***Abstract* – The feature learning algorithm for information-extreme classifier by clustering of Fast Retina Keypoint binary descriptor, calculated for local features, and usage of spatial pyramid kernel for increasing noise immunity and informativeness of feature representation are considered. Proposed a method of parameters optimization for feature extractor and decision rules based on multi-level coarse features coding using information criterion and population-based search algorithm.**

***Keywords* – coarse coding; Fast Retina Keypoint; feature extraction; machine learning; classifier; information criterion.**

I. INTRODUCTION

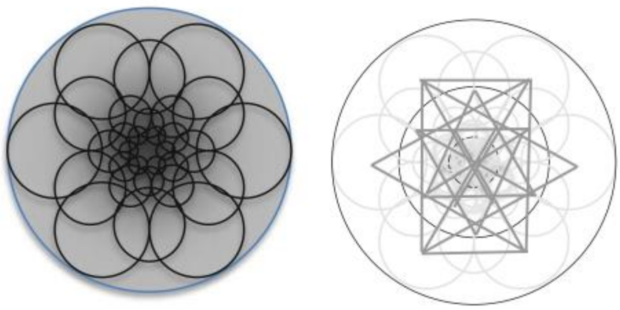
One way to improve throughput and reduce power consumption in autonomous and distributed intelligent systems is the implementation of computationally efficient algorithms. The process of intelligent processing be divided into three stages: collection and data preprocessing, feature extraction and classification or regression analysis. For a detailed analysis of data, common method of filing is the transformation of a two-dimensional matrix, which can be interpreted as an image. That is why in recent years the main direction in evolution of intelligent systems is related to research and development trainable feature extractors for visual data.

Recent experience shows that the most effective approach to the analysis of visual data is to use local features as patch descriptors that describe the variation of the brightness gradient in the neighborhood of the key points of the image [1]. The binary descriptor of key points is faster and simpler. Among them the most effective algorithm in terms of accuracy is FREAK (Fast Retina Keypoint) descriptor, which simulates the behavior of the cells of the eye. However, there is a lack of effective algorithms for descriptor quantization and sparse coding of observations described by binary descriptors. Usage of K-means algorithm of cluster analysis for quantization binary descriptor within byte representation leads to slow convergence, dependance on the choice of initial centroids and low informativeness of obtained quantization [1]. Also high dimension of obtained observations usually leads to the intersection of classes in the feature space, overfitting and the "curse of dimensionality" during parameters regularization for classifiers based on neural networks or support vector [2,3].

A promising approach of system synthesis for intelligent data analysis is using algorithms of self-organize unstructured grids, such as the method of growing neural gas, for quantization of binary descriptors, FREAK, within bit level representation and methods of classification analysis based on the usage of coarse binary encoding floating-point descriptor by multi-level system of thresholds. In this case, transformation of feature space is performed by computationally efficient comparisons. In this case, usage of logarithmic information criterion for optimization that is more effective in condition of restricted amount of training samples owing to the smoothing effect of the logarithmic function is proposed.

II. FEATURE LEARNING

In FREAK the sampling grid shown in Fig. 1, where each circle corresponds to a receptive field and its size represents the standard deviation of the Gaussian kernels applied to the underlying sampling point [1].



a b

Figure 1. Receptive fields (a) and corresponding pairs (b) in FREAK

In an attempt to reduce noise, the smoothed intensities of these receptive fields has been thresholded based on the neighborhood information. In binary FREAK descriptor used thresholded difference between pairs of the receptive field smoothed by corresponding Gaussian kernels. Calculation binary vector performed through one-bit sequence differences:

$$F = \sum_{0 \le a < N} 2^a \, sign\left(I\left(P_a^{r_1}\right) - I\left(P_a^{r_2}\right)\right) \qquad (1)$$

where N – selected descriptor size; I($I(P_a^{r_i})$) – smoothed brightness of i -th field of a - th receptive fields pair. To adapt the algorithm neural gas for cluster analysis binary descriptors on bit level it is proposed some modification. Consider the basic steps of a modified neural gas:

1. Defined two primary nodes a and b in points w_a and w_b , which corresponding to the two first input vector. Nodes a and b are connected by an edge, whose

age is zero. Errors in the nodes a and b similarly taken a zero.

2. Selects the next input vector (signal) x.
3. Determined the nearest and the second, in the sense of closeness to Hamming distance, x nodes (and denote them s_1 and s_2 respectively).
4. Age of all edges, that incidental to node, increase per unit.
5. Counter error node s_1 increase in

$$d(x \oplus w_{s_1}) = \sum_{i=1}^{N} (x_i \oplus w_{s_1}, i).$$

6. s_1 and its topological neighbors (nodes connected to it by an edge) shifting in a direction to him at a distance $\varepsilon_b \cdot d(x \oplus w_{s_1})$ and $\varepsilon_n \cdot d(x \oplus w_n)$, where $0 < \varepsilon_b \ll 1$ and $0 < \varepsilon_n \ll \varepsilon_b$ by inverting of not matching bits in randomly selected not matching bytes.
7. If s_1 and s_2 connected by an edge, its age is reset; otherwise between s_1 and s_2 creates a new edge with age equal to zero.
8. All edges in the graph with age greater than $a_{\max}$ deleted and removed nodes with no incident edges (were isolated).
9. If the number of the current iteration a multiple of the value λ (one of the parameters of the algorithm), then carried insertion a new node at the point w_r between the nodes w_q and w_f where q - the node with the highest accumulated error; f - topological neighbor q with the greatest accumulated error. An edge between f and q removed, instead it attached edges between f and r, and also between r and q. At the same time, coordinates of the vector w_r determined by copying vector f (or q) and by inverting the half of distinct with q (or f) bit in randomly selected distinct bytes. Error in nodes f and q reduced by multiplying by a constant $\alpha < 1$; error value of the new node r is initialized by a value that is equal to the value of node error q.
10. The error of all nodes is reduced by multiplying by a constant $\beta < 1$.
11. If the input data is no more or made the maximum taken number of iterations, then stop algorithm.

Encoding observations is proposed to be implemented by constructing a histogram of occurrence of quantized local features using the Average pooling method [3]. At that assign scores of local features to codewords determined by the principle softmin. To take into account of spatial information is proposed to use Spatial Pyramid Kernel [2]:

$$K = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} \left(I^l - I^{l+1} \right), \qquad (2)$$

where I^l – histogram at l resolution level, that calculated by applying the intersection function to the histograms of cells at l resolution level.

III. INFORMATION-EXTREME CLASSIFIER

The basic idea of information-extreme classifier lies in a purposeful binary encoding of quantitative features by comparing their values with corresponding thresholds to build a clear partition feature space into classes. Creation of binary training set is carried out according to the rule:

$$x_{m,i,l}^{(j)} = \begin{cases} 1, & if\ y_{i,\max}\left[1 - \dfrac{\delta_{i,l}}{\delta_{\max}}\right] \le y_{m,i}^{(j)} \le y_{i,\max}; \\ 0, & else, \end{cases} \qquad (3)$$

where $y_{i,\max}$ – the maximum value of i-th features in the training sample; δ_i – parameter of l-th receptive field for i-th feature value $\delta_{\max}$ – the maximum value of receptive field.Calculation coordinate values of binary reference vector, Calculation coordinate values of the binary vector x_m reference against which construction takes place in the space in a radial basis Hamming container classes shall according to the rule

$$x_{m,i,l} = \begin{cases} 1, & \text{if} \quad \dfrac{1}{n_m} \sum_{j=1}^{n_m} x_{m,i,l}^{(j)} > \dfrac{1}{n} \sum_{k=1}^{M} \sum_{j=1}^{n_k} x_{k,i,l}^{(j)}; \\ 0, & \text{if} \quad \text{else}. \end{cases} , \ \mathrm{i} = \overline{1,\mathrm{N}} \qquad (4)$$

At the same time, iterative process of parameter optimization $\delta_{i,l}$ performed by maximizing the averaged information criterion on set of classes during training of classifier. Its working formula has the form

$$J_m^{(k)} = \frac{K_{1,m}^{(k)} - K_{2,m}^{(k)}}{n_m(\log(2n_m + 10^{-\omega}) + \omega)} * \\ * \log\left(\frac{10^{-\omega} + n_m + [K_{1,m}^{(k)} - K_{2,m}^{(k)}]}{10^{-\omega} + n_m - [K_{1,m}^{(k)} - K_{2,m}^{(k)}]} \right) \qquad (5)$$

where $K_{1,m}^{(k)}$ – the number of events that characterize the membership of sample of class X_m^o to the container of class X_m^o at k step of machine learning; $K_{2,m}^{(k)}$ – the number of events that characterize the membership of sample of class X_c^o to the container of class X_m^o; ω – constant that regulates the sensitivity of criteria depending on accuracy characteristics.

Admissible (working) domain of the function of information criterion (5) is limited by inequalities $K_{1,m} \geq 0{,}5 \cdot n_m$ and $K_{2,m} < 0{,}5 \cdot n_m$.

IV. RESULTS

The proposed algorithms were tested on a survey data matrix of EMG sensors. Each observation was presented in the format of HD-EMG map with resolution 72x100 [5].

Training sample contains 300 observations per class. Fig. 1a shows the dependence of parameter λ and information criterion, known algorithm neural gas for quantization the byte representation of binary descriptors and fig. 1b shows the dependence of parameter λ and information criterion modified algorithm neural gas for quantization binary descriptors on bit level.

Search maximum of information criterion carried by the algorithm Particle swarm optimization in parameter space $\delta_{i,l}$. The number of agents swarm was equal to 30, no particle acceleration [6 ,7].

Analysis of Fig. 2 and Fig. 3 shows that the modified algorithm of neural gas provides a maximum value of information criteria at higher values of the parameter λ, which corresponding to a smaller number of clusters.

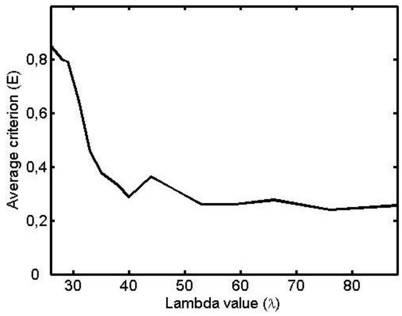


Figure 2. Results of machine learning with classic algorithm neural gas

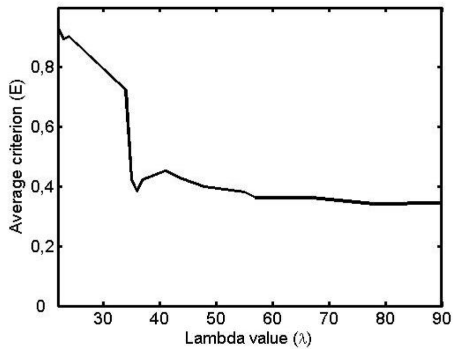


Figure 3. Results of machine learning with modified

At the same time quantization of descriptors on the bit level unlike quantization of descriptor on the byte level allowed to obtain highly accurate decision rules that are close to unmistakable.

V. CONCLUSIONS

So features learning algorithm for information-extreme classifier is the consistent implementation of transformation: construction of two-dimensional maps of signals, calculating local features, quantization of descriptors, coding observations using Spatial Pyramid Kernel and optimization of quantization parameters during the learning process of information-extreme classifier. Moreover, the proposed modification of the neural gas for quantization of binary descriptors on bit level can increase the accuracy of decision rules at a lower computational cost.

REFERENCES:

[1] Hilario Gomez C., Medathati K., Kornprobst P., Murino V., Sona D. “Improving FREAK Descriptor for Image Classification,” Lecture Notes in Computer Science, vol. 9163, pp. 14–23, June 2015 [Digests 10th International Conference, ICVS, Copenhagen, Denmark, July 6-9, 2015]

[2] Qiong Yao, Xiang Xu., “Freak Descriptor With Spatial Pyramid Kernel For Scene Categorization,” Inte. Conference on Mechatronics, Electronic, Industrial and Control Engineering Published by Atlantis Press, pp. 126–129, April 2015.

[3] S. Gao, L. Duan, and I. W. Tsang, “DEFEATnet–A Deep Conventional Image Representation for Image Classification, ” IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 3, pp. 494-505, March 2016

[4] V. Moskalenko, A. Korobov, “Information-extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extractor,” Radio Electronics, Computer Science, Control., in press. (in Ukrainian)

[5] I. S. Jacobs and C. P. Bean, “High-density surface EMG maps from upper-arm and forearm,” Journal of NeuroEngineering and Rehabilitation, vol. 60, issue 2, pp 2–17, December 2012.

[6] A. S Dovbysh, A. S Moskalenko, V. V. Moskalenko, I. V. Shelehov, “Designing algorithms for optimization of parameters of functioning of intelligent system for radionuclide myocardial diagnostics,” Eastern-European Journal of Enterprise Technologies,vol. 3, no. 3/9 (81), pp. 11-18, May 2016.

[7] V. Moskalenko, S. Pimonenko, “Optimizing the parameters of functioning of the system of management of data center IT infrastructure,” Eastern-European Journal of Enterprise Technologies, vol. 5, no. 3 (83), pp. 21-29, September 2016.