NON-PARAMETRIC CLASSIFICATION METHODS IN IMAGE RECOGNITION

Varlamova Lyudmila P.
Tashkent University of Information Technologies, 108 Amir Timur street, Tashkent, Uzbekistan,
e-mail:vlp@bk.ru

Abstract

Image recognition tasks use different approaches when it is necessary to restore the distribution function of a continuous random quantity under the condition of limiting the law of distribution. Similar situations occur when the image is noisy and there is a small sample of observations. The article discusses non-parametric methods of restoring the distribution function in recognition tasks.

Key words: image, nonparametric methods, small sampling, image recognition, random quantity distribution function.

I. INTRODUCTION

In the problems of image recognition in the conditions of small samples of observations and with uneven distribution, the presence of interference or noise, they face problems of eliminating interference and maximum recognition of image objects. The main indicators are the stability of recognition algorithms in the presence of applicative interference (image noise due to object shadowing, the presence of affected areas), when there is a small sample of observations and a priori data are not available, when applying the Laplace principle is difficult or impossible. For example, in (Duda et al.1976; Mokeev et al. 2013; Lapko et al. 1999), the problems of applying approaches to overcoming the problem of a small number of samples were considered using the methods of reducing dimensionality and adaptive nonparametric identification algorithms, and discriminant analysis methods (Mokeev et al. 2013; Lapko et al. 1999).

For example, in the works [1-3] the issues of application of approaches to overcoming the problem of small number of samples using methods of dimension reduction and adaptive algorithms of nonparametric identification are considered.

In such tasks, object recognition is an integral part of the image processing process. The result of solving the recognition problem in this situation is not only the class of the found object, but also its position. The uncertainty of the characteristics of objects makes the task of recognition in mathematical and computational terms extremely complex and time-consuming. Problems are compounded by a high degree of image distortion. With the purpose of reducing the dimension and as one of the methods of determining the characteristics, the measured intensity of grey color for each pixel is used (value 0 - corresponds to black, 255 - to white) [4].

Typically, the analysis object is a rectangular image that is converted into a vector of gray intensity values by line-by-line scanning the original image. This image presentation method has the disadvantages of redundancy and bulkiness. However, the use of parametric techniques, i.e., recognition/recovery using the characteristics of individual pixels, is the most complete, taking into account all available data, which is especially important in the initial development of image processing techniques.

The purpose of this work is to compare "window," nonparametric methods in image recognition tasks under conditions of small observation samples.

II. STATEMENT OF THE PROBLEM

One approach in image recovery is the Bayesian approach in building a complex method of image recovery when the statistical properties of image classes where the class is the image of an individual object are considered to be given. Based on this approach, it is possible to build a classification rule that is optimal, in the sense that its use provides, on average, the least loss in making classification errors. Due to the specifics of the case in question (classes of image objects have different sizes, there is a small sample), we use an estimate of the parameters of the Bayesian type [4,5]. It is assumed that a class set exists ω_j, j=1,2,…,M, distribution of a priori probability for each class of vectors of $p(\omega_j)$ depends on a vector of parameters c, distribution density of intensity of pixels of $p(x(A_j)|\omega_j)$in the A_j area, are distributed independently. An imaging function that specifies the probability distribution of ω_j images with hidden variable data $p(\omega_j \mid x)$.

$$\frac{p(x|\omega_j)p(\omega_j)}{p(x)}, \quad p(x,\omega) = p(x_1,\omega_1) \times \cdots \times p(x_n,\omega_n). \tag{1}$$

The model of a particular image is a set of values of these hidden variables, the posteriori probabilities of which can be estimated by the Bayes rule [4], in this case taking the form

$$p(x|\omega_j) \sim p(\omega_j, | x)\, p(x) \tag{2}$$

The simplest model where the luminance of individual pixels are distributed according to the same normal law characterized by two parameters a (mean) and ς (variance)

$$p(\omega_j|(a,\varsigma)) = \prod_{x\in\Omega} p(\omega_j\,(x)|(a,\varsigma)) = \coprod_{x\in\Omega} \frac{1}{\sqrt{2\pi\varsigma^2}} e^{-(\frac{x-a}{\sqrt{2\varsigma}})^2}. \tag{3}$$

An image model is represented as a sequence of pixels with different shades of gray. Each of the shades is connected with each other, the image is formed sequentially, is also restored sequentially.

The distribution function of a random variable associated with the distribution density allows you to find the desired parameter values for each pixel of the region A_{ij} of the image row j:

$$\frac{p(x|x_j)p(x_j)}{p(x)} = \frac{p(x\Big|A_j)p(x\left(\frac{A}{A_j}\right)|x_j)p(x_j)}{p\big(x(A_j)\big)p(x\left(\frac{A}{A_j}\right))} = \frac{p(x(A_j\Big|x_j)p\left(x\left(\frac{A}{A_j}\right)\right)p(x_j)}{p\big(x(A_j)\big)p(x\left(\frac{A}{A_j}\right))} = \frac{p(x(A_j)|x_j)p(x_j)}{p(x(A_j)} \quad (4)$$

A_j is the image area of a separate image element (row) c^i with a pixel intensity distribution density $p(x(A))$ of a certain region *A*.

The image in the region A_j is formed from elements of objects or as a union $A_{i_1}, \dots, A_{i_k}$, where $A_{i_1}, \dots, A_{i_k}$, but $A_j \subseteq \bigcup_{l=1}^{k} A_{i_l}$, while the number of unions is equal to the number of sets $(i_1, \dots, i_k)$. The probability of occurrence of the union $A_{i_1}, \dots, A_{i_k}$ is equal to $\prod_{l=1}^{k} p_{i_l}$ (4')

Normal distribution density is $p(x\left(A_i \cap \bigcup_{l=1}^{k} A_{i_j}\right))$ with moments of distributions $p(x(A_{i_l})|\omega_{i_l}$.

If in the region $\bigcup_{l=1}^{k} A_{i_l}$ there is no part of image or image object, then the density of the normal distribution, taking into account (4), has the form [4]:

$$p\left(x(A_j)\right) = \int p(x|\omega_j)\, p(\omega|x_1 \dots x_j) \quad (5)$$

$$p\left(x(A_j)\right) = \sum_{l=1}^{r_j} p_l (2\pi)^{-n(A_j/2} \, |C_l|^{-1/2} exp\left\{-\frac{(x-a_l)^T C_l^{-1}(x-a_l)}{2}\right\}, \quad (6)$$

r_j-is the number of different associations of image elements of the region A_j; p_l - probability of occurrence of a union (4 '); a_l, C_l are the moments of the normal distribution of the union. The Bayesian rule allows you to choose an image model based on the posterior probability $p(x|x_j)$ through the a priori probability *p(x)*. Thus, this allows the analysis of images, identifying changes and constructing velocity fields from a series of images (for example, a video sequence). Window (nuclear or mask) smoothing methods for analyzing large volumes, for estimating data distributions in the presence of a number of interference, allow smoothing and finding the distribution density of random variables when the distribution density is unknown and there is no a priori information about its parametric form. Under these conditions, it is possible to restore the function *p(x)* by nonparametric methods: histograms, nearest neighbors, and the Parzen – Rosenblatt approximations.

III. PARZEN-ROSENBLATT DISTRIBUTION DENSITY RECONSTRUCTION

The distribution density calculated using window (nuclear) functions is described by expression (6) of the form

$$\hat{p}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right), \quad (7)$$

where n – sample size; K is the nuclear (window) function; h is the window width; x is a random sample; x_i is the *i*-th implementation of a random variable.

In the multidimensional case, the density estimate, taking into account (4) and (4 ')

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{m}\frac{1}{h^j}K\left(\frac{x^j - x_i^j}{h^j}\right), \quad (8)$$

where m – space size, *kernel* - a function used to restore the distribution density, a continuous bounded function with a unit integral

$$\int K(y)dy = 1, \quad (9)$$

$$\int y\, K(y)dy = 0,$$

$$\int y^i K(y)dy = k_i(K) < \infty.$$

Function (9) with properties $K(y) \geq 0, is\ K(y) = K(-y)$. The kernel is a non-negative bounded symmetric real function whose integral is equal to unity, the statistical moments must be infinite. The order ν of function (9) is equal to the order of the first moment, which is not equal to zero. If $k_1(K) = 0$ and $k_2(K) > 0$, then K is a second-order kernel (ν = 2).

Known nuclear functions of the second order (Fig. 1):

- Epanechnikov kernel $K(y) = \frac{3}{4}(1 - y^2)$;
- Gauss kernel $K(y) = \frac{1}{\sqrt{2\pi}}e^{-0.5y^2}$;
- Laplass kernel $K(y) = \frac{1}{2}e^{-|y|}$;
- uniform kernel $K(y) = \frac{1}{2}, |y| \leq 1$;
- triangular kernel $K(y) = 1 - |y|, |y| \leq 1$
- biquadratic kernel $k(y) = \frac{3(1-y^2)}{4}, \ |y| \leq 1$

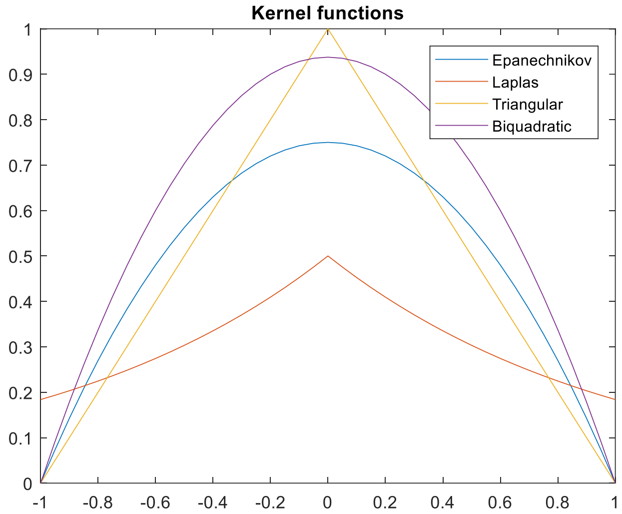

Fig. 1 Examples of various kernel functions

The optimal values of the nuclear function and parameter h are found from the condition that the functional $J = \int ln\, K(y) \cdot K(y)dy$ reaches the maximum value.

Or in other words: to restore the empirical distribution density using the Parzen-Rosenblatt window, the unknown parameter is the window width h in expression (7). Therefore, to determine the empirical density, it is necessary to solve the problem of finding the window width, so as to find the optimal h_{opt}. Finding the optimal window size is made from the condition

$$\left(\frac{R(K)}{k_2^4(K)S_D(a(h_{opt}))}\right)^{1/5} n^{1/5} - h_{opt} = 0. \quad (10)$$

The Parzen – Rosenblatt method allows one to construct an approximation of the distribution function of any finite random sequence, which, provided the parameter h is optimized, turns out to be quite smooth [7,8].

The search for the optimal window width can be carried out by other methods. The accuracy of the restored dependence depends little on the choice of the kernel. The kernel determines the degree of smoothness of the function.

Using the Parzen-Rosenblatt method, an approximation of the distribution function of a random sequence with a limited scattering region was constructed

$$F(x; x_0, \sigma, l) = \int_{x_{min}}^{x} f\,(\xi; x_0, \sigma, l)d\xi, \quad (11)$$

where $f_{lim}(x; x_0, \sigma, l) = K\left[\phi(x; x_0, \sigma, l) + \sum_{n=0}^{\infty} \phi_{2n+1}^{\pm}\,(x; x_0, \sigma, l) + \sum_{n=1}^{\infty} \phi_{2n}^{\pm}\,(x; x_0, \sigma, l)\right]$,

x_0- the position of the scattering center in the coordinate system with the origin in the center of the segment $[x_{min}, x_{max}]$,

σ- standard deviation (SD) of a random function in the absence of restrictions,

$l = x_{max} - x_{min}$ – span scatter,

K– normalization coefficient [9],

$x_{2n+1}^{\pm}, x_{2n}^{\pm}$ determined by the formulas:

$x_{2n}^{\pm} = \pm 4nl + x_0,\ x_{2n+1}^{\pm} = \pm(4n+2)l - x_0,$

When analyzing the quality of the approximation of the distribution function of a random sequence by the Parzen-Roseblatt method and the k – nearest neighbors (k-NN) method in the scattering region [-5; 5], with the standard deviation of the random variable σ = [1,3,5,7,10], the results were obtained shown in table 1 and 2. Graphs of convergence are shown in Fig. 2 and 3.

IV. RESEARCH METHODOLOGY

Table 1. The error of the estimation of the distribution function by the Parzen-Rosenblatt method

SD \ Diapazon	-5	-3	0	3	5
1	0,001432	0,00078	0,0001389	0,00079	0,001428
3	0,000227	0,00023	0,00008821	0,000398	0,000553
5	0,000279	0,00022	0,0001638	0,00018	0,000201
7	0,0002	0,000181	0,0001298	0,000125	0,000152
10	0,000143	0,000138	0,0001379	0,000161	0,000147

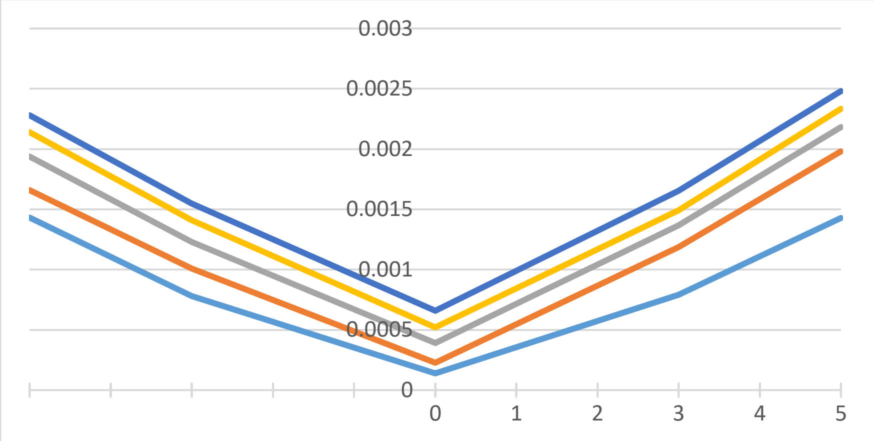

Fig. 2. Restoration of the distribution density function by the Parzen – Rosenblatt method

Table 2. Error of estimation of distribution function by k – nearest neighbors method

SD \ Diapazon	-5	-3	0	3	5
1	0,0006412	0,00004523	0,00001498	0,00004267	0,0001125
3	0,00007934	0,00005424	0,00002274	0,00005917	0,00009254
5	0,00002315	0,00002884	0,00004868	0,00005254	0,00004132
7	0,00003157	0,00005793	0,00002141	0,00006232	0,00006778
10	0,00005682	0,0006147	0,00002798	0,00001356	0,00001067

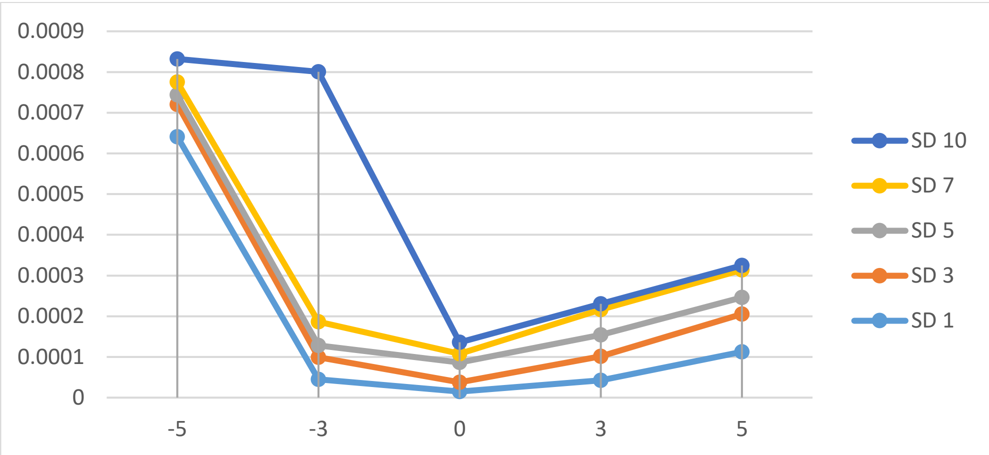

Fig. 3. Recovery of the density function by the method of k-NN

The use of the approximation of the distribution function of a random variable by the methods of Parzen - Rosenblatt and *k* – NN method with different nuclear functions showed the relative proximity of the approximating function and the true distribution function. In the literature [10-12], there are a number of works with analytical data on the issue of comparing the Parzen – Rosenblatt method with imaginary sources, histograms

V. CONCLUSION

In this paper, the nonparametric Parzen – Rosenblatt and *k* – NN methods were considered. For a small sample size, the matrix of two-dimensional distribution parameters becomes singular, and for small window widths this method reduces to the *k*-NN method [13], which has its own characteristics such as dependence on the selected step and instability to errors. In this case, it becomes necessary to impose conditions on the distribution density$p_{y,h}(x)$, the function and the width of the window. Accordingly, the amount of data [2] in the image set is growing. However, a similar problem can be solved by methods of reducing the dimension or by methods of discriminant analysis [2,14,15]. Moreover, to reduce the volume of the data set, an external image database is used. Then the task of constructing a classifier [16-18] is greatly simplified, and the problem with a minimum sample and the least number of standards is reduced to a problem with a minimum sample.

The Parzen – Rosenblatt method showed good convergence or, if the value of the blurriness parameter h is chosen correctly, guarantees smooth estimates of the distribution function.

REFERENCES

[1] Duda R., Hart P. Pattern recognition and scene analysis. Translation from English. G. G. Vajeshtejnv, A. M. Vaskovski, edited by V. L. Stefanyuk, “MIR” Publishing House, Moskow 1976. – 509 p.

[2] Mokeev V.V., Tomilov S.V. On the solution of the problem of small sample size when using linear discriminant analysis in face recognition problems // Business Informatics. – 2013. №1(23). Pp..37-43.

[3] Lapko A.V., Chencov S.V., Lapko V.A. Non-parametric patterns of pattern recognition in small samples // Autometry. -1999. -№6. pp. 105-113.

[4] Wapnik V.N., Chervonenkis A.J. Pattern recognition theory (statistical learning problems). Publishing House "Science," Main Edition of Physical and Mathematical Literature, Moscow, 1974, -416p.

[5] Rafael C.Gozalez, Richard E.Woods Digital Image Processing. Publisher: Pearson; 4 edition (March 30, 2017) (4th Edition). ASIN: 0133356728. 1192 pages.

[6] Voronin_ V.V. Methods and algorithms of image recovery in conditions of incomplete a priori information: monograph / V.V.Voronin_ V.I.Marchuk. –Shahti _ GOU VPO «YuRGUES»_ 2010. _89p.

[7] Rozenblatt M. Remarks on some nonparametric estimates of density function// Ann. Math. Statist., 1956, 27, 832–835 p.

[8] Parzen E. On estimation of probability density function and mode // Ann. Math. Statist., 33, 3, 162, 1065–1076 p.

[9] Porshnev S.V._ Koposov A.S. Ob osobennostyah chislennih ocenok parametra razmitosti plotnostei raspredelenii sluchainih posledovatelnostei v approksimacii Rozenblatta_Parzena // V mire nauchnih otkritiy. – 2013. http://ej.kubagro.ru/2013/08/pdf/76.pdf.

[10] Katkovnik, V. Nonparametric density estimation with adaptive varying window size. – Signal Processing Laboratory, Tampere University of Technology. – 2000. http://www2.mdanderson.org/app/ilya/ Publications/europtoparzen.pdf.

[11] Izenman, A. J. Recent developments in nonparametric density estimation / A. J. Izenman // J. Am. Statistical Assoc. – 1991. – Vol. 86. – P. 205–224.

[12] Jeon, B. Fast parzen density estimation using clustering-based branch and bound / B. Jeon, D. Landgrebe // IEEE Trans. Pattern Analysis and Machine Intelligence. –1994. – Vol. 16, no. 9. – P. 950–954.

[13] Omachi, S. A fast algorithm for a k-NN classifier based on branch and bound method and computational quantity estimation. http://citeseer.ist.psu.edu/611715.html.

[14] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, MaS, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Li FF. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 2015; 115(3): 211-252. DOI: 10.1007/s11263-015-0816-y.

[15] Girolami, M. Probability density estimation from optimally condensed data samples / M. Girolami, C. He // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2003. – Vol. 25, no. 10. – P. 1253–1264.

[16] Tan, X. Face recognition from a single image per person: a survey / X. Tan, S. Chen, Z.H. Zhou, F. Zhang // Pattern Recognition.– 2006.– Vol. 39, Issue 9.– P. 1725-1745. – DOI: 10.1016/j.patcog.2006.03.013.

[17] Bertinetto, L. Learning feed-forward one-shot learners / L. Bertinetto, J.F. Henriques, J. Valmadre, P. Torr, A. Vedaldi // Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016. – 2016. – P. 523-531.

[18] Aidu F. A . , Vapnik V . N. Estimating the probability density by the stochastic Regularization method// Automation and Remote Control., 1989, Issue 4, 84–97. http://www.mathnet.ru/