Experimental Studies on the Recognition of Small-Sized Objects in Video Images Using Multidimensional Spatial-Subband Vectors

Vera A. Goloshapova, Evgeniy G. Zhilyakov, Aleksandr N. Zalivin, Evgeniy M. Mamatov, Ivan I. Oleynik

Belgorod State University, Pobedy Str., 85, Belgorod, 308015, Russia, E-mail: oleinik_i@bsu.edu.ru

Abstract

Background: A decision rule has been developed to recognize small-sized objects in video images. The input data for the decision rule are the samples of space-subband vectors formed from the image of objects. Experimental studies of the decisive function are carried out using images of various small-sized objects. They demonstrate obtained numerical values of the like-lihood ratio logarithm used to make a decision on object recognition. It is shown that the devel-oped decision rule allows to recognition of small-sized objects in video images when carrying out a priori training.

Methods: The use of sub-band representation of video images allows to keep both the spatial and frequency structure of the object image. This can give an advantage over other methods using the recognition rules with a preliminary description and the parameters of the recognition feature informativeness. It is possible to use the methods of subband analysis and signal syn-thesis.

Results: The studies were carried out using images of unmanned aerial vehicles of the copter type and showed an image of an object according to which a training sample (a reference im-age) and a stud-ied image were formed. Experimental studies showed that the largest values of the likelihood ratio logarithm were located in proportion to those pixels of the image under study, on which the object is located.

Conclusion: The produced decision rule authorizes to recognition of various small-sized ob-jects in video images with high-quality indicators. The invented approach to the construction of the decision rule makes it possible to use optimal solutions and the Neumann-Pearson criterion to set the threshold level.

Keywords: Decision function, Assessment, Recognition, Experimental research, Vector, Covariance ma-trix, Subband.

1. Introduction

Deep learning has had a profound impact on various branches of technology over the past few years. One of the hottest topics in this industry is computer vision [1]. The field of image pro-cessing is constantly expanding and growing. During the last five years, there has been a signif-icant increase in image processing and its possibilities such as image morphology, neural net-works, color image processing, image data compression, image recognition and image analysis systems.

Image is better than any other form of data for understanding by humans [2]. Sight allows hu-mans to understand the world around them. Image understanding, image analysis and computer vision aims to repeat and imitate the effect of human vision through computer understanding and image understanding, digital image processing requires an operation whose steps are clear and fixed and applied to each pixel of the image. In the first stage of image processing, this op-eration is applied pixel by pixel [3, 4].

Detection and recognition of objects in video images, including small-sized ones, requires solv-ing a number of interrelated problems. There are many methods and algorithms based on the comparison of images using various approaches and mathematical models. The following methods are used for the recognition of fiducial marks, integral, contour, and characteristic points, etc. [5]. The method of characteristic points is simple to implement, and has high per-formance, but does not provide invariance to the shift and rotation of the image object. When using integral methods, computational costs are increased, since the position parameters are calculated from the information about the entire grayscale image.

The contour method is based on the use of fiducial contours. In this case, the recognition is car-ried out using the degree of similarity of the found contour and standards. The degree of simi-larity is determined by the correlation coefficient [6]. To implement this method, it is neces-sary to isolate and analyze a contour with high noise immunity. Here, the volume of computa-tional costs remains quite high, although it is decreased with integral methods [7].

The algorithms for moving object detection, often used in CCTV cameras, are well-developed. However, such algorithms do not recognize the objects and do not detect stationary objects [8, 9]. For example, this manifests itself in the detection of small-sized unmanned aerial vehicles (like a copter, for example), which can remain stationary for a long time.

In the general case, to solve the recognition problem, it is necessary to select informative fea-tures of recognition objects, select a method for describing the features that minimize computa-tional costs for recognition, and select a decision-making procedure in real time [10]. The spec-ificity of the object recognition problem is a significant a priori uncertainty regarding the num-ber of object classes, their features, and characteristics, which does not allow to use of tradi-tional methods of object recognition, focused on post-processing data.

2. Spatial subband representation of video images

The use of sub-band representation of video images allows you to preserve both the spatial and frequency structure of the object image [11]. This can give an advantage over other methods using the recognition rules with a preliminary definition and the parameters of the recognition feature informativeness. For this, it is possible to use the methods of subband analysis and sig-nal synthesis [11, 12].

A video image can be represented as the pixels with bit representation (for example, P = 2c, c – bits) of brightness levels. The number of bits determines the number of gradations of bright-ness levels from black to white, and one line of the image (formed by the pixels of the selected image fragment) can be represented in vector form [13].

$$\vec{\mathbf{S}} = \left(s_1, \quad s_2, \quad \ldots \quad s_Q,\right), \; q = 1,\ldots,Q \quad (1)$$

Where Q, q, and s are the number of pixels in a line, pixel number, and the number of pixel gradation levels (brightness) respectively.

In general, the image column formed by pixels can be similarly represented in vector form. Then we will consider row-by-row representations, and the change in the pixel brightness of the original image (signal components) will correspond to some "spatial frequencies".

In the general case, the energies of the signal components are concentrated in a small number of rather narrow intervals of the spectrum definition range. With this approach, it is possible to split the frequency axis into several frequency intervals [12].

$$\Delta\omega = 4\pi / (Q-1). \qquad (2)$$

According to relation (2), it is possible to divide the frequency axis into frequency intervals (subbands) in the following form [12].

$$\Omega_k = [-\Omega_{1k}, -\Omega_{2k}) \cup [\Omega_{1k}, \Omega_{2k}), \quad k = 0, ..., K \qquad (3)$$

$$\Omega_{10} = 0; \; \Omega_{20} = 2\pi / (Q-1); \; \Omega_{1k} = \Omega_{2k-1}; \; \Omega_{2k} - \Omega_{1k} = 4\pi / (Q-1), \qquad (4)$$

Where K is the number of frequency intervals. It is customary to use the concept of a part of the signal energy on the basis of subband analysis, falling into a given frequency interval [12].

$$P_k(s) = \int_{\omega \in \Omega_k} |S(\omega)|^2 \, d\omega / 2\pi \qquad (5)$$

The representation of the signal in the area of originals can be written in a quadratic form [12].

$$U_k(\vec{s}) = \vec{s}^T \mathbf{A}_k \vec{s}, \qquad (6)$$

Where A_k is the subband matrix with the elements.

$$a_{\gamma\xi}^k = \frac{\sin[\Omega_{2k}(\gamma - \xi)] - \sin[\Omega_{1k}(\gamma - \xi)]}{\pi(\gamma - \xi)} \text{ at } \gamma \neq \xi;$$

$$a_{\gamma\xi}^k = \frac{\Omega_{2k} - \Omega_{1k}}{\pi} \text{ at } \gamma = \xi; \qquad \gamma, \xi = 1, ..., Q.$$

The subband matrix is calculated for each frequency interval k. After transformation (6), the vector (1) is actually transformed into the vector of dimension k, which can be written in the following form [12, 13].

$$\vec{\mathbf{U}} = \begin{pmatrix} U_{(1)} & U_{(2)} & \dots & U_{(k)} \end{pmatrix}^T, \qquad (7)$$

Where U is the fraction (the part) of the signal energy in the frequency range; k = 1,…,K – the frequency interval number, and T is the transposition sign.

The vector of the form (7) can be called a space-subband vector (SSV).

An image with the dimension N x Q pixels can be represented as a sample of vectors of the volume N with the dimension K.

$$\mathbf{U}^{(N)} = \begin{pmatrix} U_{(1)1} & U_{(1)2} & \dots & U_{(1)N} \\ U_{(2)1} & U_{(1)2} & \dots & U_{(1)N} \\ \dots & \dots & \dots & \\ U_{(k)1} & U_{(k)2} & \dots & U_{(k)N} \end{pmatrix}, \qquad (8)$$

Where N is the sample size, the first character (in brackets) at U denotes the frequency interval number, and the second symbol at U (without brackets) denotes the number of the vector in the sample i=1… N.

In this case, N will denote the number of image lines, and K is the number of frequency intervals (subbands). In the general case, the transformation can also be carried out by columns (then N will denote the number of columns).

In this case, the vector is a multidimensional random variable. Its dimension is equal to the number of frequency intervals (subbands). In fact, it is the space-subband vector (SSV). The image formation process is subject to random disturbances, the probabilistic nature of which affects all stages. This assumption makes it possible to use a statistical approach to obtain the estimates of the SSV sample distribution. The probability distribution of the sample can be characterized by two moments, the first initial (mathematical expectation) and the second cen-tral (covariance matrix) [14]. Estimation of mathematical expectation (ME) vector of the sam-ple (8) is determined by the expression [14]

$$\vec{\mathbf{m}}_{(k)} = \frac{1}{N}\sum_{i=1}^{N}\vec{\mathbf{U}}_{(k)i} . \qquad (9)$$

The dimension of the vector makes K.

The estimate of the covariance matrix is calculated in accordance with the expression [14].

$$\mathbf{M}_{(k\times k)} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}\right)\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}\right)^{T} . \qquad (10)$$

The elements of the covariance matrix reflect the degree of statistical connection between the elements of the original vector of the fixed parameters. In fact, it is a subband covariance ma-trix (SCM) with the dimension .

With a Gaussian distribution of the sample, the probability density of the ith sample vector can be written in the form of the expression [15]

$$P\left(\vec{\mathbf{U}}_{(k)i}\right) = \frac{1}{(2\pi)^{n}\left(\det \mathbf{M}_{(k\times k)}\right)^{n}} \exp\left[-\frac{1}{2}\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}\right)^{T}\mathbf{M}^{-1}_{(k\times k)}\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}\right)\right]. \qquad (11)$$

Such a presentation of data when processing information allows you to apply a statistical ap-proach and describe the data obtained from the video image of an object in the form of a Gauss-ian distribution with an estimate of the spectral covariance matrix of the object.

3. Decision rule of recognition

The property of feature normality greatly simplifies the form of the decision function, since the decision function turns out to be a linear combination of observations, and its distribution will be normal again [14]. All types of decision rules are based on the formation of the likelihood ratio and its comparison with a certain threshold, the value of which is determined by the se-lected quality criterion [14].

In statistical recognition, the probability densities of feature distribution are unknown a priori, therefore, not the probability densities are substituted into the decision rules, but their esti-mates obtained during the learning process. Accordingly, not the likelihood ratio is compared with the threshold in the decisive rule, but its estimate obtained in the course of training [14, 17].

In this case, the task of parametric learning will be to estimate the parameters (the vector of means and the SCM) of the normal probability densities used in the decision rule. Therefore, we will operate not with distribution moments, but with their estimates.

The original vector has the dimension k, which determines the number of frequency intervals (or sub-bands).

Consider the construction of decision rules using the example of small unmanned aerial vehicle (UAV) recognition. An image of the UAV is shown in Figure 1. Let us formulate hypotheses.

There is no object (UAV) in the image (the hypothesis). There is an object in the image (the hypothesis). In this case, it is necessary to use the sample obtained a priori for the object as a training sample. The sample obtained during the current estimation of the parameters of the studied image is used as a control sample. In this case, the recognition has two alternatives. This makes it possible to use fairly well-known criteria. For example, the threshold can be de-termined by the Neumann-Pearson criterion [14, 16], setting the probability of a type I error (false alarm). The smaller the set value , the greater the threshold. The decision rule (the log-arithm of the likelihood ratio) can be written as [14, 15].

$$\mathrm{L}=\frac{n}{2}\ln\frac{\left|\mathbf{M}_{(k\times k)}^{(1)}\right|}{\left|\mathbf{M}_{(k\times k)}^{(0)}\right|}+\frac{1}{2}\cdot\sum_{i=1}^{n}\Big(\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}^{(1)}\right)^{\mathrm{T}}\cdot(\mathbf{M}_{(k\times k)}^{(1)})^{-1}\cdot\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}^{(1)}\right)-$$

$$-\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}^{(0)}\right)^{\mathrm{T}}\cdot(\mathbf{M}_{(k\times k)}^{(0)})^{-1}\cdot\left(\vec{\mathbf{U}}_{(k)i}-\vec{\mathbf{m}}_{(k)}^{(0)}\right)\Big)>\ln C\,, \qquad (12)$$

Where the superscripts with the estimates m and M denote the hypothesis (1 corresponds to the hypothesis , and 0 refers to the hypothesis), and subscripts determine the dimension.

Since the objects on the image are small (i.e., the size of the object is significantly smaller than the image itself), the algorithm should provide for the current estimation of the control sample parameters. The size of the "window" in which the estimation is carried out is chosen commen-surate with the object size. This "window" must go through the entire image (for example, line by line and by columns, with a shift by a certain number of pixels). After each current assess-ment of the control sample parameters, it is necessary to calculate the logarithm of the likeli-hood ratio (12) and compare it with a given threshold.

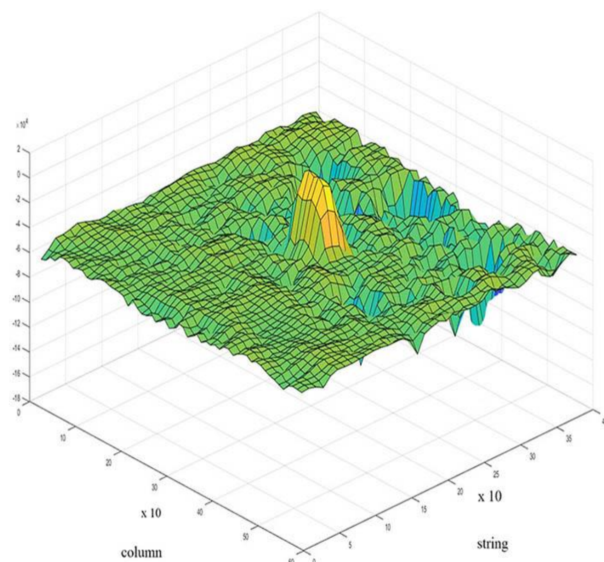
This decision rule allows you to recognize a specific object (a small-sized UAV) according to which the training was carried out.

4. Experimental studies

The studies were carried out using images of unmanned aerial vehicles of the copter type. Fig-ure 1 shows an image of an object according to which a training sample (a reference image) and a studied image were formed. Figure 2 shows an image, which demonstrates the value of the likelihood ratio logarithm (12), in proportion to the rows and columns of the studied image pixels.

Figure 1 Reference and studied images

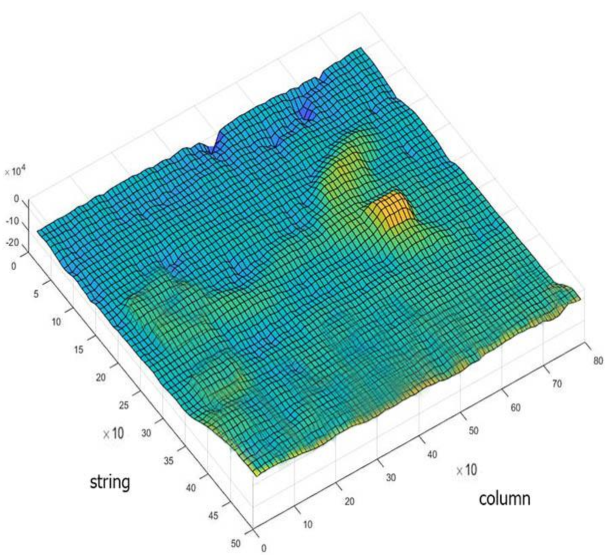
Figure 2 Values of likelihood ratio logarithm



Similarly, experimental studies of the image with two different objects (copters) were carried out. At the same time, the training sample was formed for one object only. Figure 3 shows the object image, according to which the training sample (reference image) and the studied image were formed. Figure 4 shows an image, which demonstrates the values of the likelihood ratio logarithm (12), in proportion to the rows and columns of the pixels of the image under study.

Figure 3 Reference and studied images

Figure 4 Values of likelihood ratio logarithm



Experimental studies show that the largest values of the likelihood ratio logarithm are located in proportion to those pixels of the image under study, on which the object is located. Accord-ing to this object the training was carried out.

5. Discussion

The developed decision rule uses the Bayesian approach during statistical hypothesis testing. The decision rule is to calculate the decision function and compare it with a threshold. At the same time, the experimental data are presented in the form of a sample of vectors and are de-scribed by multidimensional probability densities with Gaussian distributions [16]. The esti-mates are the mathematical expectation estimates for the samples of spatial-subband vectors of the input data and their covariance matrices. When testing the hypothesis about the presence of an object on an image, it is possible to use the Neumann - Pearson criterion [15, 17]. To demonstrate the efficiency of the decision rule, experimental studies were carried out using field data (video images). Experimental studies have shown that the developed decision rule makes it possible to recognize small-sized objects in video images, when carrying out a priori training.

6. Conclusions

1. The developed decision rule allows to recognition of various small-sized objects in video images with high-quality indicators.

2. The developed approach to the construction of the decision rule makes it possible to use op-timal solutions and the Neumann-Pearson criterion to set the threshold level.

3. Experimental studies using the field data confirm the possibilities of the developed decision rule for small object recognition in video images.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Computational intelligence and neuroscience. 2018 Feb 1.
2. Kulkarni G, Premraj V, Ordonez V, Dhar S, Li S, Choi Y, Berg AC, Berg TL. Babytalk: Understanding and generating simple image descriptions. IEEE

transactions on pattern analysis and machine intelligence. 2013 May 31; 35(12):2891-903.
3. Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. Cengage Learning; 2014 Jan 21.
4. Moeslund TB, Hilton A, Krüger V. A survey of advances in vision-based human motion capture and analysis. Computer vision and image understanding. 2006 Nov 1; 104(2-3):90-126.
5. Furman YaA. Introduction to contour analysis and its application to image and signal processing. FIZMATLIT. 2004, 456.
6. Girenko AV, Lyashenko VV, Mashtalir VP, Putyatin EP. Methods of correlation detection of objects. Kharkiv, JSC "Businessinform". 1996, 112.
7. Vasin YuG, Lebedev LI. Recognition of composite image objects based on structural and correlation-extreme methods. Collection of reports of the 13th All-Russian Conference "Mathematical Methods of Pattern Recognition" (MMRO-13), Zelenogorsk. 2007, 148: 285-288.
8. Ko T. A survey on behavior analysis in video surveillance for homeland security applications. In2008 37th IEEE Applied Imagery Pattern Recognition Workshop 2008 Oct 15 (pp. 1-8). IEEE.
9. Chen Z, Khemmar R, Decoux B, Atahouet A, Ertaud JY. Real time object detection, tracking, and distance and motion estimation based on deep learning: Application to smart mobility. In2019 Eighth International Conference on Emerging Security Technologies (EST) 2019 Jul 22 (pp. 1-6). IEEE.
10. Jurie F, Triggs B. Creating efficient codebooks for visual recognition. InTenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 2005 Oct 17 (Vol. 1, pp. 604-610). IEEE.
11. Zhilyakov EG, Belov SP, Oleinik II, Babarinov SL. Trubitsyna, D.I. Generalized sub band analysis and signal synthesis. Bulletin of Electrical Engineering and Informatics. 2020, 1(9): 78-86.
12. Zhilyakov EG. Optimal subband methods for analyzing and synthesizing signals of finite duration, Automation and Telemechanics. 2015, 4: 51-66.
13. Zhilyakov EG, Belov SP, Oleinik II, Prokhorenko EI. Regularization of Inverse Signal Recovery Problems, HELIX the Scientific Explorer. 2019, 9 (2): 4883- 4889.
14. Fomin YaA, Tarlovsky GR. Statistical theory of pattern recognition. Moscow, Radio and Communication. 1986, 264.
15. Burdanova EV, Zhilyakov EG, Mamatov AV, Nemtsev AN, Oleynik II. Decisive rule experimental studies to detect objects on the background of the earth surface using polarization differences of radar signals. COMPUSOFT. An International Journal of Advanced Computer Technology, 2019, 8(6): 3166-3170.
16. Goloschapova VA, Kalashnikov PA, Oleynik II. Parametric decision rules for object recognition in multi-dimensional vector representation of color images. Science, Educatition, Practice: materials of the International University Science Forum (Canada, Toronto), 2020 May 27: 238-246.
17. Zalivin AN, Oleynik II, Pirogenko YA. Decision rule for recognizing small objects based on subband processing of radar signals. Science, Educatition, Practice: materials of the International University Science Forum (Canada, Toronto), 2020 April 22: 197-207.