Computer simulation of adapting the object image selection method to video surveillance conditions

N P Kornyshev and D A Serebrjakov
Yaroslav-the-Wise Novgorod State University, Veliky Novgorod, Russian Federation
E-mail: Nikolai.Kornishev@novsu.ru

Abstract. The article deals with the issues of computer modeling of methods for selecting images of objects against a non-uniform background. A test video sequence with given background and object parameters is considered, which provides imitation of one of the special cases of video surveillance conditions, namely, the convergence of the video surveillance point and the object. The issues of adaptation of the compensation selection method to the specified conditions of video surveillance are discussed. Examples of test images and graphs of dependences of the probability of correct determination of coordinates depending on the value of the local contrast of the object in relation to the background, obtained by computer simulation are given.

1. Introduction
A fairly large number of works [1-5] are devoted to the methods of selecting the image of an object against an uneven background, including taking into account the changing visible dimensions of the object during the approach. Nevertheless, in the design of television and optoelectronic systems intended for the detection and tracking of targets, a very critical parameter is the speed of the system and the ease of implementation. From this point of view, it is of practical interest to increase the efficiency of using compensation selection methods based on background subtraction, such as, for example, the method of selecting an image of a small-sized object against a non-uniform background, described in the literature [5].

This method consists in processing the current image signal by an anti-aliasing or median filter with a mask size m larger than the object image size $M_o$, but much smaller than the size of the images of background fragments $M_f$, i.e.: $m > M_o$, but $m << M_f$ (1). In practice, a square $m \times m$ mask with $m \leq 9$ elements is usually used. Compensation is made by subtracting the filtered signal from the original current signal. The output binary signal of the object is formed by comparing the difference signal with a fixed threshold.

This article discusses the adaptation of the above method to the conditions of approaching an object with a video observation point in order to increase its efficiency by additional processing of the input signal, which ensures the exchange of resolution for signal-to-noise ratio as the size of the object increases.

2. Results and discussion
Increasing the size of the object image leads to non-observance of conditions (1), and, accordingly, to the termination of suppression of the object signal during filtering. As a result, the object signal is compensated along with the background during subsequent subtraction.
If a smoothing filter is used, then if conditions (1) are not met, the subtraction of the smoothed signal from the current one leads to the formation of the contours of the object using the "unsharp masking" method [6]. Despite the fact that the presence of object contours is sufficient to determine, in particular, the coordinates of its center, the amplitude of the object contours formed during "unsharp masking" turns out to be less than the amplitude of the object signal formed when condition (1) is met. Thus, a decrease in the signal-to-noise ratio that appears in this case leads to a decrease in the selection accuracy of the object image.

Note also that if a median filter is used, then after subtraction, if condition (1) is not met, the contour elements of the object image are also formed. However, depending on its configuration, the resulting contour can have multiple breaks. So, for example, instead of a contour image of a square, only points corresponding to the corners of the image will be formed. Moreover, depending on the orientation of the square, the number and location of such points can change, which can lead to significant errors in determining the coordinates of the center of the object image. From this point of view, as well as from the point of view of the simplicity of the hardware implementation, the use of an anti-aliasing filter is preferable. Therefore, in what follows we consider the adaptation of this method taking into account the use of a smoothing filter in it.

Adaptation to video surveillance conditions consists in evaluating the area $S_{\text{b}}$ of the image of the selected object in each $i$-th frame and enlarging the video signal reading aperture to mixmi elements in the case of a corresponding increase in $S_{\text{b}}$. Initially, the size of the image of an object, a detected object in the video surveillance field, is one pixel, that is $M_i=1$. The current size $M_{\text{i}}$ of the object is determined based on the concept of the equivalent scattering area (ESR), namely $M_{\text{oi}} = \sqrt{S_{\text{oi}}}$.

As the sizes $M_{\text{i}}$ of the object image in the $i$-th frame increase, the size $m_i$ of the video signal reading aperture is taken equal to an integer part of the number $m_i=M_{\text{i}}$, within the values $1 < m_i \leq m$. As a result of this enlargement of the aperture, the signal-to-noise ratio also increases by a factor of $m_i$, while simultaneously ensuring condition (1). When the equality $m_i=m_\text{i}$ is reached, the size of the reading aperture is fixed. So, for example, when the aperture size $m_i=9$ is reached, the further formation of the object signal will be carried out already at the signal-to-noise ratio 9 times higher than the initial one.

The resulting decrease in the resolution does not affect the result of the selection of the image of the object due to the corresponding increase in its size when the object approaches the video observation point. The smoothing of the object signal front that occurs in this case also does not affect the result of determining the coordinates, since the center of the object will correspond to the maximum of its image signal. Thus, the technical result of the adaptation is to improve the accuracy of the object image selection [7].

Computer simulation of the operation of the selection device considered above was carried out in the MATLAB environment according to the method described in [8]. In this case, we used a test video sequence [8] with a variable value of the object contrast $K$ in relation to the maximum background value from 1% to 10% with a step of 1% with a signal-to-noise ratio $SN=6$ and a mask size of $m=7$. An example of simulating an object and a non-uniform background in a test video sequence is shown in figure 1.

The object image size was increased by dilating the logical signal $F = \text{imdilate} (F, c)$ with a periodic (after a given number of frames) increasing the size of the mask $c = \text{ones} (z0, z0)$ with a step of 1, starting from one value. The imitation of the change in the position of the object in the frame was carried out by periodically performing the standard operation of rotating the image $F$ by a given angle $R0$ with "filling" the corners formed in the image with a black (zero) background: $F = \text{imrotate} (F, R0, \text{"cro"}).$The generation of the background image with a given irregularity and with a given dimension of fragments with smooth boundaries was carried out by generating a sinusoidal signal $T = U_{\text{max}} \sin (wt + \phi)$ in each line of the image frame. The resulting image matrix $T$ was transposed and added to the original matrix. Thus, the magnitude of the inhomogeneity was determined by the amplitude of the sinusoidal signal, $2U_{\text{max}}$, and the dimension of the background fragments, by the
frequency \( w \). The background movement was simulated by autoincrementing the phase shift \( \phi \) in each frame of the video sequence.

The output parameter in the simulation was the probability \( p(t) \) of the coincidence in each frame of the coordinates of the selected object with the tolerance field. In turn, the tolerance field was specified in each frame as the neighborhood of the center of the reference object within the filter mask size \( m \).

The output parameter \( p(t) \) was defined as the ratio of the number of frames of the video sequence, in which the coordinates of the selected object coincided with the tolerance field, to the total number of frames of the video sequence.

Based on the simulation results, the dependences \( p(t) = f(K) \) at \( SN = \text{const} \) were found. In addition, the average, maximum and standard deviation of the coordinates of the selected object from the reference values was estimated for all frames of the test video sequence. Figure 2 shows the plots of \( p(t) = f(K) \) dependencies obtained in the process of computer simulation of the selection method with adaptation [7] to a change in the object image size (graph 2) and without adaptation [5] (graph 1).

Figure 2. Dependency plots \( p(t) = f(K) \) with adaptation of the method (row 2) and without adaptation (row 1) at \( SN = 6, m = 7 \).

Figure 2 shows that adaptation to the conditions of increasing the size of the object's image increases the likelihood of coincidence in each frame of the coordinates of the selected object with the tolerance field, and, consequently, the accuracy of the object selection when the contrast decreases with respect to the background.
Figure 3. Comparative diagrams of the mean (mean dX, mean dY), maximum (max dX, max dY), and root-mean-square (RMS, dX, RMSdY) deviations of the coordinates of the selected object from the reference values for the adapted (row 2) and initial (row 1) method.

Figure 3 shows comparative diagrams of the mean, maximum and standard deviation of the coordinates of the selected object from the reference values for all frames of the test video sequence for the adapted and the original method. The diagrams show that adaptation significantly reduces the deviations of the above parameters, which indicates an increase in the accuracy of the object image selection.

3. Conclusion
When the image of an object is enlarged in the process of its approach to the point of video surveillance, adaptation to the conditions of video surveillance provides an increase in the accuracy of the selection of images of objects. A possible increase in the size of the video signal readout aperture and a corresponding increase in the signal-to-noise ratio is determined by the size of the spatial filter mask used in the background compensation procedure.

References
[1] Tsytsulin A K, Pavlov V A, Bobrovsky A I and Morozov A V 2019 Information estimates in the problems of detection-estimation-transmission of a signal in space television Questions of radioelectronics. Ser. Televisiontechnology 3 61–74
[2] Tsytsulin A K, Bobrovsky A I, Morozov A V, Pavlov V A and Galeeva M A 2019 Application of convolutional neural networks for automatic selection of small-sized artificial space objects in optical images of the starry sky Optical journal 86 (10) 30–38
[3] Levko G V, Bobrovsky A I, Morozov A V and Tsytsulin A K 2016 Objects detection against a starry background Problems of Radioelectronics. Ser. Television technology 2 29–38
[4] Tsytsulin A K, Pavlov V A, Bobrovsky A I, Morozov A V and Zubakin I A 2019 Adaptive coding of images separated into a dominant object and a background Questions of radio electronics. Ser. Television technology 3 75–86
[5] Andreev A L, Korotaev V V and Pashkovsky A M 2013 Selection of images of small-sized objects against a non-uniform background in conditions of interference Izv. Universities. Instrumentation 56 (10) 88–93
[6] Pratt W 1982 Digital image processing: trans. from English (M. Mir) p 575
[7] Kornyshev N P and Serebryakov D A 2020 A method for selecting an image of an object against a non-uniform background Patent RU No. 2734655
[8] Kornyshev N P and Serebryakov D A 2020 Computer modeling of the target background environment in assessing the effectiveness of methods for selecting objects on a non-uniform background Vestn. Novg. state university. Ser. Technical science 2 (118) 79–83