Deformation Field Estimate for Image Sequence by Applying Stochastic Adaptation in the Block Method

Roman Kovalenko
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
r.kovalenko.o@gmail.com

Pavel Smirnov
Ventra
Moscow, Russia
rtcis@mail.ru

Radik Ibragimov
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
ibragimow.it@gmail.com

Alexander Tashlinskiy
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
tag@ulstu.ru

***Abstract*—The paper researches the block method based on stochastic adaptation, which is used to estimate the deformation field of the image sequence. The similarity model was selected as the deformation model. The method was implemented for two target functions: the mean square inter-frame difference and the inter-frame correlation coefficient. The result of the proposed method was compared with the Motion Vector Field Adaptive Search Technique. The proposed method has a high noise resistance and allows one to reduce the influence of global inter-frame geometric changes.**

Keywords*—*stochastic adaptation, mean square difference, correlation coefficient, image sequence, block method, deformation field.

I. Introduction

Detection of the area of a moving object is usually used in machine vision systems for highlighting the areas of interest in images and subsequent analysis improvement. The task of detecting a moving object for complex cases has not yet received a general solution. The complexity of this task is caused by the possibility of various dynamic changes in the scene (smooth, sharp or local changes in lighting conditions, weather changes, repetitive movement, etc.). A more complex case can be observed when the background is similar to a moving object. Therefore, the development of algorithms analyzing scene movement in difficult conditions remains a relevant subject.

The task of detecting the area of a moving object is considered as the task of dividing image pixels into two groups: background and foreground, where the foreground is the moving object. The foreground may consist of one or several objects. In both cases, the foreground objects must be detected, and if there are several objects, the moving objects must also be separated from each other.

As with many other image processing tasks, moving object detection can be implemented in both spatial and frequency domains.

In the frequency domain, most of the moving object detection methods are based on wavelet transformations [1] and low order fractional statistics [2]. Background changes have less effect on the result of the moving object area detection in the frequency domain than in the spatial domain. But with this approach, problems with shadows appear [3]. Methods in the frequency domain have high computational complexity, so in practice, they are used much less often than methods of the spatial domain.

There are different approaches to detect the area of a moving object in the spatial domain: methods based on inter-frame difference estimation[4, 5], background subtraction [4, 6], statistical [5, 7], block method [8], optical flow analysis [9, 10]. In this paper, we develop an algorithm based on a block method.

II. Problem Statement

Most methods for the deformation field $\mathbf{H}$ estimation use inter-frame image processing. In this case, the image of a moving object can be represented as some region or regions of the current image having inter-frame geometric changes (IGC). Thus dividing the image into nonoverlapping areas (blocks) and estimation their inter-frame deformation parameters, the deformation field $\mathbf{H}$ will be obtained. The obtained field is used to determine image blocks that correspond to a moving object, e.g. by using a threshold. This approach corresponds to the general principle of block methods for detecting motion [11], which are based on finding the corresponding location of blocks of the current (deformed) frame on the previous (reference) frame. To do this, the current frame $\mathbf{Z}^t$ of the image sequence is divided into many nonoverlapping blocks $B_{i,j}$, where i, j is the block center coordinates. The size of blocks is selected based on the size of objects whose movement needs to be detected. The solution comes down to finding the motion vector $\bar{h}_{i,j}$ of each block $B_{i,j}$ on frame $\mathbf{Z}^{t-1}$.

$$\bar{h}_{i,j} = \arg\left(\underset{v_{i,j}\in O}{\text{extremum}}\left(Q\left(i, j, v_{i,j}\right)\right)\right) \quad (1)$$

where O – is the search area, $Q(i, j, v_{i,j})$ – is the target function of matching blocks of the current and the previous frames. By assigning the shift $\bar{h}_{i,j}$ to the nodes of the reference grid included in block $B_{i,j}$, we obtain the deformation field $\mathbf{H} = \{\bar{h}_{i,j}\}$ for the deformed image and the reference image. This approach provides high efficiency at a relatively low computational complexity [8, 11].

Block methods assume static background on which moving objects are to be detected. In practice, consecutive frames can have global mutual spatial deformations, e.g. due to camera movements. In this case the algorithm based on the block method will detect motion in almost the entire frame. To solve this problem, a more complex models for determining the location of blocks $B_{i,j}$ such as similarity model [12] can be chosen. This models include the following parameters $\overline{\alpha}^{t,t-1} = (\overline{h}, \varphi, \kappa)^T$: shift along the basic axes $\overline{h} = (h_x, h_y)^T$, rotation angle φ and scale κ . The paper proposes to estimate the location of blocks $B_{i,j}$ by stochastic adaptation procedure [13] to find the parameters of $\overline{\alpha}^{t,t-1}$. The algorithm is resistant to impulse noise and requires small computational cost which is virtually independent of block sizes. Block sizes are usually significantly smaller than the size of the object to be detected.

III. Algorithm decription

For each block $B_{i,j}$ of the reference frame, the stochastic block method proposes a recurring finding of estimation parameters (vector $\overline{\alpha}_{i,j}^{t,t-1}$) position on the deformed frame in accordance with the procedure [13]:

$$\hat{\overline{\alpha}}_{i,j(n)}^{t,t-1} = \hat{\overline{\alpha}}_{i,j(n-1)}^{t,t-1} - \mathbf{\Lambda}_n \overline{\beta}_n (J(\hat{\overline{\alpha}}_{i,j(n-1)}^{t,t-1}, Z_n)) \qquad (2)$$

where $\overline{\beta}$ – stochastic gradient of the target function $J(\cdot)$; $\mathbf{\Lambda}_n$ –the array of learning rate; Z_n – a local sample, it used to find $\overline{\beta}$ at the iteration, $n = \overline{0, N-1}$; N – the number of iterations. Note that a local sample Z_n is independently selected for each estimation iteration.

The method was implemented for two most common target functions: the mean square inter-frame difference (MSID) and the inter-frame correlation coefficient [14]. When using MSID for the stochastic gradient at the n-th iteration, we obtain [15]:

$$\beta_{in} = \frac{1}{2\mu\Delta x}\sum_{l=1}^{\mu}\left(\Delta\tilde{z}_x^t\left(\tilde{z}_{xl+\Delta x,yl}^t + \tilde{z}_{xl-\Delta x,yl}^t - 2z_{il,jl}^{t-1}\right)\right)\frac{\partial x}{\partial\alpha_i} + \\ + \frac{1}{2\mu\Delta y}\sum_{l=1}^{\mu}\left(\Delta\tilde{z}_y^t\left(\tilde{z}_{xl,yl+\Delta y}^t + \tilde{z}_{xl,yl-\Delta y}^t - 2z_{il,jl}^{t-1}\right)\right)\frac{\partial y}{\partial\alpha_i}, \qquad (3)$$

where (x_l, y_l) – coordinates on image $\mathbf{Z}^t$; (i_l, j_l) – coordinates on image $\mathbf{Z}^{t-1}$; $\tilde{z}_{xl,yl}^t$ is the brightness of the oversampling image $\mathbf{Z}^t$ taking into account the estimates $\hat{\overline{\alpha}}_{i,j(n-1)}^{t,t-1}$, obtained in the previous iteration; $\Delta x, \Delta y$ the steps of finding derivatives $\partial\tilde{z}_{xl,yl}^t/\partial x$ and $\partial\tilde{z}_{xl,yl}^t/\partial y$ using the finite difference [14], μ is local sample size Z_n . Partial derivatives $\partial x/\partial\overline{\alpha}$ and $\partial y/\partial\overline{\alpha}$ are found analytically.

When inter-frame correlation coefficient is used as the target function, then the expression of the stochastic gradient on the n-th iteration takes the form:

$$\beta_{in} = \frac{1}{2\mu\hat{\sigma}_{t-1}}\left[\sum_{l=1}^{\mu}\left(\frac{z_{il,jl}^{t-1} - z_m^{t-1}}{\Delta x}\left(\frac{\tilde{z}_{xl+\Delta x,yl}^t}{\sigma_{+\Delta x}} - \frac{\tilde{z}_{xl-\Delta x,yl}^t}{\sigma_{-\Delta x}}\right)\right)\frac{\partial x}{\partial\alpha_i} + \right. \\ \left. + \sum_{l=1}^{\mu}\left(\frac{z_{il,jl}^{t-1} - z_m^{t-1}}{\Delta y}\left(\frac{\tilde{z}_{xl,yl+\Delta y}^t}{\sigma_{+\Delta y}} - \frac{\tilde{z}_{xl,yl-\Delta y}^t}{\sigma_{-\Delta y}}\right)\right)\frac{\partial y}{\partial\alpha_i}\right], \qquad (4)$$

where $\sigma_{\pm\Delta x}^2 = (\mu-1)^{-1}\left(\sum_{l=1}^{\mu}\left(\tilde{z}_{xl\pm\Delta x,yl}^t\right)^2 - \mu\left(\tilde{z}_{\pm\Delta xm}^t\right)^2\right)$, $\hat{\sigma}_{t-1}^2 = (\mu-1)^{-1}\sum_{l=1}^{\mu}\left(z_{il,jl}^{t-1} - z_m^{t-1}\right)^2$; $\tilde{z}_{\pm\Delta xm}^t$ and $z_m^{t-1} = \mu^{-1}\sum_{l=1}^{\mu} z_{il,jl}^{t-1}$ – the mean values of $\tilde{z}_{xl\pm\Delta x,yl}^t$ and $z_{il,jl}^{t-1}$.

The method based on MSID requires less computational costs and can work already with the local sample size $\mu = 1$, which allows it to be implemented in pixel-by-pixel processing. Therefore, in the proposed method, the choice of MSID as the main target function is appropriate.

If the similarity model is used as a model for geometric deformations of the reference and deformed frame, then the derivatives $\partial x/\partial\alpha_i$ and $\partial y/\partial\alpha_i$ will be defined by expressions:

$$\partial x/\partial h_x = 1 ,$$

$$\partial x/\partial h_y = 0 ,$$

$$\partial x/\partial\kappa = (a_l - x_o)\cos\varphi - (b_l - y_o)\sin\varphi ,$$

$$\partial x/\partial\varphi = -\kappa((a_l - x_o)\sin\varphi + (b_l - y_o)\cos\varphi) ,$$

$$\partial y/\partial h_x = 0 ,$$

$$\partial y/\partial h_y = 1 ,$$

$$\partial y/\partial\kappa = (a_l - x_o)\sin\varphi + (b_l - y_o)\cos\varphi ,$$

$$\partial y/\partial\varphi = \kappa((a_l - x_o)\cos\varphi - (b_l - y_o)\sin\varphi) ,$$

where (x_o, y_o) - coordinates of the rotation center.

Usually to represent deformation field, every reference pixel coordinates (x, y) is set in accordance with the shift vector $\overline{h} = (h_x, h_y)^T$. To obtain such deformation field representation, the estimates of the deformation parameters $\hat{\overline{\alpha}}_{i,j}$ must be recalculated using the accepted deformation model. In particular, for the similarity model, we get:

$$\hat{h}_{(i,j)x} = x_o + \hat{\kappa}_{n-1}((i - x_o)\cos\hat{\varphi}_{n-1} - (j - y_o)\sin\hat{\varphi}_{n-1}) + \hat{h}_{x(n-1)} , \qquad (5)$$

$$\hat{h}_{(i,j)y} = y_o + \hat{\kappa}_{n-1}((i - x_o)\sin\hat{\varphi}_{n-1} + (j - y_o)\cos\hat{\varphi}_{n-1}) + \hat{h}_{y(n-1)} . \qquad (6)$$

The algorithm can be described simplified way as follows. For neighboring frames that do not have mutual global IGC, parameter estimates of blocks without motion will remain close to zero in contrast to blocks with motion, whose parameter estimates will converge to some nonzero values (for a scale $\kappa = 1$). Described rule is a criterion for assigning a block to motion. If neighboring frames have mutual global IGC, then the estimates of the deformation parameters of all blocks will be different from zero. In this case, the blocks corresponding to the moving object will form compact clusters. Blocks with global deformations are located throughout the frame, which is used as a criterion for determining global deformations [16]. The deformation parameters of moving objects are determined by subtracting the global deformations.

IV. Experimental results

Fig. 1 shows an example of two consecutive frames of the image sequence $\mathbf{Z}^t$, which was obtained with a microscope at a magnification of 400 times. On this figure, you can see two unicellular Sonderia organisms. An organism that is completely in the frame is in motion. Motion parameters can be written by using the similarity model: $\bar{h} = (3.6, -2.1)^T$, $\varphi = 3^\circ$, $\kappa = 1$. And the second organism is almost motionless. At the same time frames have global IGC with parameters: $\bar{h} = (1, -2.2)^T$, $\varphi = -1^\circ$, $\kappa = 1.01$. Also for a complex case, unbiased additive Gaussian noise with a signal/noise ratio of 14 dB was added to the images.

Fig. 1. An example of an image sequence.

Fig. 2 shows the comparative results of the inter-frame difference algorithms Fig. 2(a), background subtraction Fig. 2(b) and the proposed stochastic block method Fig. 2(c). For ease of comparison, each image has an organism contour.

(a) (b) (c)

Fig. 2. The result of motion detection by different algorithms.

Fig. 2 shows that the inter-frame difference and background subtraction algorithms define the second organism in motion, due to global geometric changes in consecutive frames. These two algorithms detect an area of a moving object with a large number of gaps, especially in low-contrast places where there is a small gradient of image brightness. The proposed stochastic block method highlights a region of motion with almost no gaps. The gaps can only correspond to blocks in which most of the pixels relate to the background and only some of them relate to a moving object.

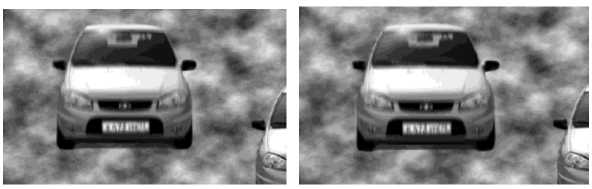


Fig. 3. Example of an image sequence with a moving object.

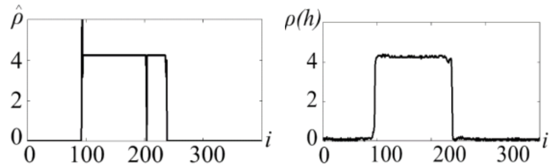
As already noted, the proposed method also works for pixel-by-pixel estimation of the deformation field. In this case, each element of the deformation field contains information about the direction and magnitude of the pixel shift in the reference image relative to its position on the deformed image. For example, Fig. 3 shows two consecutive frames of an image sequence in which the car in the center is moving and the car on the right is stationary. At the same time images of a moving car have the following parameters of inter-frame spatial shift: $h_x = 3$, $h_y = 2.95$.

The results of estimating the deformation field using the proposed method in comparison with the results obtained using a well-known blocks method named Motion Vector Field Adaptive Search Technique (MVFAST) [17] are shown below. MVFAST also allows pixel-by-pixel estimate of the deformation field. In this case, the estimates $\hat{h}_{(i,j)x}$, $\hat{h}_{(i,j)y}$ are recalculated into the vector module and its angle:

$$\rho(h) = \sqrt{\left(\hat{h}_{(i,j)x}\right)^2 + \left(\hat{h}_{(i,j)y}\right)^2}, \tag{7}$$

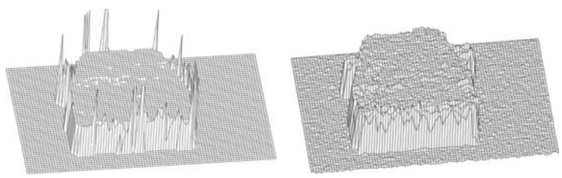
$$\varphi(h) = arctg\ \left(\hat{h}_{(i,j)x} / \hat{h}_{(i,j)y}\right). \tag{8}$$

Fig. 4 shows typical shift estimations of image pixels corresponding to the nodes for one row of the reference image. Here Fig. 4(a) corresponds to the application of MVFAST method, Fig. 4(b) to the proposed method. For MVFAST method in contrast to the proposed one, you can see the errors on the borders of the object image and in the areas inside. Gaps inside the object occur in low-contrast areas. The proposed method due to the inertia of changes in the estimates does not have this disadvantage.



(a) (b)

Fig. 4. Example of shift estimates for row.



(a) (b)

Fig. 5. Deformation field visualization.

Table 1 shows the estimation for the expected value $\hat{m}$ and variance $\hat{D}$ for both the row and the entire image. Also, the table shows that the estimation expected value of the MVFAST method for the motion area are several times greater (about 5 times for a row, 8 times for an image) than for the proposed method. The variance estimation for the motion area in the MVFAST method is many times greater than the variance of the proposed method. For a motionless area, the MVFAST method shows slightly better results for the entire image in the absence of noise. Deformation field estimates for the entire image are shown in Fig. 5: Fig. 5(a) when using the MVFAST method, Fig. 5(b) the proposed method. The Fig.5 shows significant errors in the MVFAST method at the boundaries of the object, as well as in low-contrast areas within the object.

TABLE I. THE ESTIMATION ERRORS OF SHIFT VECTORS

Algorithm	Motion area		Motionless area	
	$\hat{m}$	$\hat{D}$	$\hat{m}$	$\hat{D}$
One line processing results				
Proposed algorithm	0.01	26	0.02	3
MVFAST	0.05	2530	0.01	1
Average results of the entire image				
Proposed algorithm	0.01	140	0.09	4
MVFAST	0.08	1860	0.02	5

V. CONCLUSION

The developed method, based on identificationless stochastic adaptation, has high noise immunity and allows one to get rid of the influence of global IGC, as well as to remove small moving objects that are not of interest. In this paper, such objects were small organisms and particles, in other situations it can be rain, snow, falling leaves, etc. The detection of small objects is realized by reducing the size of blocks, up to one pixel.

ACKNOWLEDGMENT

The work was supported by RFBR and Government of Ulyanovsk Region according to the research projects № 18-41-730011 and 19-29-09048.

REFERENCES

[1] B. Antic, V. Crnojevic and D. Culibrk, “Efficient wavelet based detection of moving objects,” 16th International Conference on Digital Signal Processing, 2009.

[2] A.M. Bagci, Y. Yardimci and A. Çetin, “Moving object detection using adaptive subband decomposition and fractional lower-order statistics in video sequences,” Signal Processing, vol. 82, no. 12, pp. 1941-1947, 2002.

[3] B.U. Töreyin, A.E. Çetin, A. Aksay and M.B. Akhan, “Moving object detection in wavelet compressed video,” Signal Processing: Image Communication, vol. 20, no. 3, pp. 255-264, 2005.

[4] S. Ahmed, K. El-Sayed and S. Elhabian, “Moving object detection in spatial domain using background removal techniques,” Recent Patents on Computer Sciencee, vol. 1, no. 1, pp. 32-54, 2008.

[5] B. Karasulu and S. Korukoglu, “Moving object detection and tracking in videos,” Performance Evaluation Software SpringerBriefs in Computer Science, pp. 7-30, 2013.

[6] L. Wang and N. Yung, “Extraction of moving objects from their background based on multiple adaptive thresholds and boundary evaluation,” IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 1, pp. 40-51, 2010.

[7] R.V. Kutsov and A.P. Trifonov, “Detection of a moving object in the image,” Journal of Computer and Systems Sciences International, vol. 45, no. 3, pp. 459-468, 2006.

[8] S.V. Grishin, D.S. Vatolin, A.S. Lukin, S.Iu. Putilin and K.N. Strelnikov, “A review of block-based methods for estimating motion in digital video signals,” Software systems and tools: Thematic collection, vol. 9, pp. 50-62, 2008.

[9] N.Iu. Zolotykh, V.D. Kustikova and I.B. Meerov, “An overview of the methods for searching and tracking vehicles on the video stream,” Vestnik of the Nizhny Novgorod University. N.I. Lobachevsky, vol. 5, no. 2, pp. 348-358, 2012.

[10] H. Chen, S. Ye, A. Nedzvedz, O. Nedzvedz, H. Lv and S. Ablameyko, “Traffic extreme situations detection in video sequences based on integral optical flow,” Computer Optics, vol. 43, no. 4, pp. 647-652, 2019. DOI: 10.18287/2412-6179-2019-43-4-647-652.

[11] I.S. Zaqout, “An efficient block-based algorithm for hair removal in dermoscopic images,” Computer Optics, vol. 41, no. 4, pp. 521-527, 2017. DOI: 10.18287/2412-6179-2017-41-4-521-527.

[12] D. Pons and Zh. Forsait, “Computer vision. Modern approach,” Moscow: Viliams, 2004.

[13] A.G. Tashlinskii, “Estimation of spatial deformation parameters of image sequences,” Ulyanovsk: ULSTU, 2000.

[14] A.G. Tashlinskii, P.V. Smirnov and L.S. Biktimirov, “Methods of finding gradient estimates of target function for measurement of images parameters,” Pattern Recognition and Image Analysis, vol. 21, no. 2, pp. 339-342, 2011.

[15] A.G. Tashlinskii, P.V. Smirnov and S.S. Zhukov, “Analysis of methods of estimating objective function gradient during recurrent measurements of image parameters,” Pattern Recognition and Image Analysis, vol. 22, no. 3, pp. 399-405, 2012.

[16] A.G. Tashlinskii, S.V. Voronov and P.V. Smirnov, “A way to predict parameters of image registration by estimating inter-frame deformation of local fragments,” Pattern Recognition and Image Analysis, vol. 24, no. 1, pp. 179-184, 2014.

[17] P.I. Hosur and K.K Ma, “Motion vector field adaptive fast motion estimation,” Second International Conference on Information, Communications and Signal Processing, pp. 7-10, 1999.