

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Formation of Inter-Frame Deformation Field of Images Using Reverse Stochastic Gradient Estimation

*Alexander Tashlinskii and Pavel Smirnov*

## Abstract

The effectiveness of the use of stochastic gradient estimation to detect motion in a sequence of images is investigated. Pixel-by-pixel reverse estimation of the deformation field is used. The representation of the shift vector is considered both by the projections on the basic axes and by the polar parameters. Two approaches to estimate the parameters of the deformation field are proposed and analyzed. In the first approach, the stochastic gradient procedure sequentially processes all rows of an image to find estimates of shifts for all points of the reference image. The joint processing of the results allows compensating the inertia of the stochastic estimation. In the second approach, to improve the accuracy of estimation, the correlation of image rows is taken into account. As a criterion for the formation of the resulting estimate, the minimum of gradient estimation and correlation maximum were investigated. The computational complexity of the proposed algorithms is investigated. The algorithms are compared with the MVFAST algorithm. Examples of experimental results on the formation of the deformation field, the selection of a moving object area, and the finding of the movement and trajectory parameters of a moving object in a video sequence are given.

**Keywords:** image processing, video sequence, motion, shift vector, detection, stochastic estimation, stochastic gradient, trajectory, reverse estimation, correlation

## 1. Introduction

One of the challenges in video processing is moving object detection and tracking. Some tasks require only detection of the motion, while others require extraction of the moving object or the motion area boundary. The biggest challenge is to estimate trajectory parameters of a moving object in a video sequence. A solution quality to the problem largely depends on the accuracy of moving object area detection, since all the information needed to determine motion parameters and trajectory of the object is extracted from the image.

There are various approaches to identify the area of moving object based on the inter-frame difference [1, 2], background subtraction [2, 3], the use of statistics [2, 4], block estimation [5], and optical flow analysis [6]. The processing can be presented as estimation of inter-frame geometric deformations of two images, one

of which can be considered as the reference image  $\mathbf{Z}^s = \{z_{i,j}^s\}$  and the second as deformed image  $\mathbf{Z}^d = \{z_{i,j}^d\}$ , where  $z_{i,j}^*$  is the value in node  $(i,j)$  of the image. The field  $\mathbf{H} = \{\bar{h}_{i,j}\}$  of inter-frame shift vectors  $\bar{h}_{i,j}$  of all points of the reference image corresponding to the nodes of the sample grid will be called deformation field or shift vectors' field.

An important task of estimating the trajectory of a moving object is to increase the spatial accuracy of detecting the area of motion in an image sequence. The solution of this problem depends on the quality of estimation of the deformation field. When finding estimates of shifts for each node of the deformed image, the approach of using stochastic gradient estimation [7–9] is promising. It finds coordinates  $(x,y)$  of points of the reference image on the deformed image. That is, it evaluates vectors  $\bar{h}_{i,j} = (h_{(i,j)x}, h_{(i,j)y})^T$  of inter-frame shifts of points  $(i,j)$  (**Figure 1**) of the reference image.

In view of the insignificant change in brightness on adjacent frames of a video sequence, it is advisable to use mean square inter-frame difference as the objective function of estimation. Then, for the projections,  $h_{(i,j)x}$  and  $h_{(i,j)y}$  of shift vector, we get [10]

$$\hat{\bar{h}}_{i,j+1} = \hat{\bar{h}}_{i,j} - \lambda_h \text{sign} \left( \bar{\beta}_h \left( z_{i,j+1}^d, \hat{\bar{h}}_{i,j} \right) \right), i = \overline{1, N_x}, j = \overline{1, N_y - 1}, \quad (1)$$

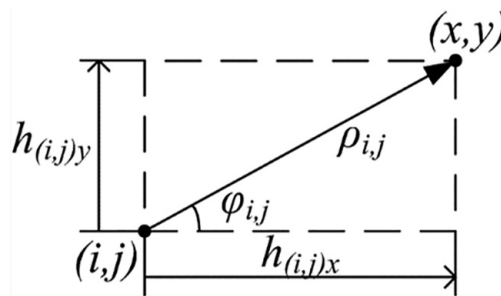
where  $\lambda_h$  is the learning rate, which determines the rate of change of the estimated parameters,  $\bar{\beta}_h = (\beta_{hx}, \beta_{hy})^T$  is the gradient estimation of an objective function, and  $N_x \times N_y$  is the image size.

The projections of stochastic gradient can be represented as [11]

$$\begin{aligned} \beta_{hx} &= \left( \tilde{z}_{x+\Delta x, y}^s - \tilde{z}_{x-\Delta x, y}^s \right) \left( \tilde{z}_{x, y}^s - z_{i,j}^d \right), \\ \beta_{hy} &= \left( \tilde{z}_{x, y+\Delta y}^s - \tilde{z}_{x, y-\Delta y}^s \right) \left( \tilde{z}_{x, y}^s - z_{i,j}^d \right), \end{aligned}$$

where  $(x,y)$  are the coordinates of the reference image point on the deformed image,  $\tilde{z}_{x,y}^s$  is the brightness of the continuous image  $\tilde{\mathbf{Z}}^s$  obtained from  $\mathbf{Z}^s$  by means of interpolation, and  $\Delta x, \Delta y$  are the steps of finding estimates of derivatives  $d\tilde{z}_{x,y}^s/dx$  and  $d\tilde{z}_{x,y}^s/dy$  [12].

Note that inter-frame shift vector  $\bar{h}_{i,j}$  can be also represented in polar form:  $\bar{h}_{i,j} = (\rho_{i,j}, \varphi_{i,j})$  (**Figure 1**), where  $\rho_{i,j}$  is the length of the vector and  $\varphi_{i,j}$  is the angle with respect to the  $x$  axis (**Figure 1**). Functionally, these representations are equivalent. However, due to the inertia of estimates of shift vectors' field  $\mathbf{H}$ , when using



**Figure 1.**  
Representation of shift of reference image point  $(i,j)$ .

stochastic gradient descent, the use of parameters  $\rho_{i,j}$  and  $\varphi_{i,j}$  does not give equivalent estimates  $\hat{h}_{i,j}$  compared with the use of  $h_{(i,j)x}$  and  $h_{(i,j)y}$ . This is due to the fact that the sets of parameters have different physical meanings [13]. The answer to the question of which set is preferable for solving the problem of moving object area detection is not obvious and requires research.

## 2. Variants of estimation algorithms

Let us compare the efficiency of moving object area identification using the sets of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$ . The following method for estimating the shift vectors' field  $\mathbf{H}$  is used. At the first stage, stochastic gradient descent algorithm sequentially processes all nodes of the reference image row by row with the set of parameters  $(h_x, h_y)$ . It processes each row  $i$  bidirectionally [14], first, from the left to the right, getting, according to Eq. (1), the estimates

$$\begin{aligned}\hat{h}_{(i,j+1)x}^l &= \hat{h}_{(i,j)x}^l - \lambda_h \text{sign}\left(\beta_{hx}\left(z_{i,j+1}^d, \hat{h}_{i,j}^l\right)\right), \hat{h}_{(i,1)x}^l = \hat{h}_{(i-1,1)x}^l, \\ \hat{h}_{(i,j+1)y}^l &= \hat{h}_{(i,j)y}^l - \lambda_h \text{sign}\left(\beta_{hy}\left(z_{i,j+1}^d, \hat{h}_{i,j}^l\right)\right), \hat{h}_{(i,1)y}^l = \hat{h}_{(i-1,1)y}^l,\end{aligned}$$

and then from the right to the left getting the estimates

$$\begin{aligned}\hat{h}_{(i,N_y-j)x}^r &= \hat{h}_{(i,N_y-j+1)x}^r - \lambda_h \text{sign}\left(\beta_{hx}\left(z_{i,N_y-j}^d, \hat{h}_{i,N_y-j+1}^r\right)\right), \hat{h}_{(i,N_y)x}^r = \hat{h}_{(i-1,N_y)x}^r, \\ \hat{h}_{(i,N_y-j)y}^r &= \hat{h}_{(i,N_y-j+1)y}^r - \lambda_h \text{sign}\left(\beta_{hy}\left(z_{i,N_y-j}^d, \hat{h}_{i,N_y-j+1}^r\right)\right), \hat{h}_{(i,N_y)y}^r = \hat{h}_{(i-1,N_y)y}^r,\end{aligned}$$

where parameter  $\lambda_h$  is determined by the maximum speed of moving objects.

When using the set of parameters  $(\rho, \varphi)$ , considering that  $\partial x / \partial \rho = \cos \varphi$ ,  $\partial x / \partial \varphi = -\rho \sin \varphi$ ,  $\partial y / \partial \rho = \sin \varphi$ , and  $\partial y / \partial \varphi = \rho \cos \varphi$ , for gradient estimation  $\bar{\beta} = (\beta_\rho, \beta_\varphi)^T$ , we get [11]

$$\begin{aligned}\beta_\rho &= \left(\tilde{z}_{x+\Delta x,y}^s - \tilde{z}_{x-\Delta x,y}^s\right) \left(\tilde{z}_{x+\Delta x,y}^s + \tilde{z}_{x-\Delta x,y}^s - 2z_{i,j}^d\right) \cos \varphi \\ &\quad + \left(\tilde{z}_{x,y+\Delta y}^s - \tilde{z}_{x,y-\Delta y}^s\right) \left(\tilde{z}_{x,y+\Delta y}^s + \tilde{z}_{x,y-\Delta y}^s - 2z_{i,j}^d\right) \sin \varphi, \\ \beta_\varphi &= -\left(\tilde{z}_{x+\Delta x,y}^s - \tilde{z}_{x-\Delta x,y}^s\right) \left(\tilde{z}_{x+\Delta x,y}^s + \tilde{z}_{x-\Delta x,y}^s - 2z_{i,j}^d\right) \rho \sin \varphi \\ &\quad + \left(\tilde{z}_{x,y+\Delta y}^s - \tilde{z}_{x,y-\Delta y}^s\right) \left(\tilde{z}_{x,y+\Delta y}^s + \tilde{z}_{x,y-\Delta y}^s - 2z_{i,j}^d\right) \rho \cos \varphi.\end{aligned}$$

Then according to Eq. (1), we get the estimates

$$\begin{aligned}\hat{\rho}_{i,j+1}^l &= \hat{\rho}_{i,j}^l - \lambda_\rho \text{sign}\left(\beta_\rho\left(z_{i,j+1}^d, \hat{\rho}_{i,j}, \hat{\varphi}_{i,j}\right)\right), \\ \hat{\varphi}_{i,j+1}^l &= \hat{\varphi}_{i,j}^l - \lambda_\varphi \text{sign}\left(\beta_\varphi\left(z_{i,j+1}^d, \hat{\rho}_{i,j}, \hat{\varphi}_{i,j}\right)\right), \\ \hat{\rho}_{i,N_y-j}^r &= \hat{\rho}_{i,N_y-j+1}^r - \lambda_\rho \text{sign}\left(\beta_\rho\left(z_{i,N_y-j}^d, \hat{\rho}_{i,N_y-j+1}, \hat{\varphi}_{i,N_y-j+1}\right)\right), \\ \hat{\varphi}_{i,N_y-j}^r &= \hat{\varphi}_{i,N_y-j+1}^r - \lambda_\varphi \text{sign}\left(\beta_\varphi\left(z_{i,N_y-j}^d, \hat{\rho}_{i,N_y-j+1}, \hat{\varphi}_{i,N_y-j+1}\right)\right),\end{aligned}$$

where the parameters  $\lambda_\rho$  and  $\lambda_\varphi$  are also determined by the maximum speed of moving objects.

At the second stage, for each node  $(i, j)$ , the estimates  $\hat{h}_{(i,j)x}^l$  and  $\hat{h}_{(i,j)x}^r$  are processed jointly [15]. The optimal value of  $\hat{h}_{(i,j)x}$  is found between the estimates with some step  $\Delta_h$ . Step  $\Delta_h$  is determined by the required accuracy. If the absolute difference between the estimates  $\hat{h}_{(i,j)x}^l$  and  $\hat{h}_{(i,j)x}^r$  is less than  $\Delta_h$ , then the estimate  $\hat{h}_{(i,j)x}$  of the deformation parameter is assumed to be equal  $\hat{h}_{(i,j)x}^l$ . Otherwise, the set of possible values of the deformation parameter estimate is given by

$$\hat{h}_{(i,j)x}^m = \hat{h}_{(i,j)x}^l + m\Delta_h, m = \overline{0, k+1}, k = \left\lceil \frac{\hat{h}_{(i,j)x}^r - \hat{h}_{(i,j)x}^l}{\Delta_h} \right\rceil.$$

The optimal value of  $\hat{h}_{(i,j)x}$  from the resulting set is found using some criterion.

## 2.1 Criteria for the formation of the resulting estimation

Two criteria for the formation of the resulting estimation are studied [16]:

- Gradient estimation minimum

$$\min_{m=0, k+1} \beta_h \left( \hat{h}_{(i,j)x}^l + m\Delta_h \right), \quad (2)$$

- Correlation coefficient maximum (CC)

$$\max_{m=0, k+1} CC \left\{ \tilde{z}_{x(m)+p, y(m)+s}^s, z_{i+p, j+s}^d \right\}, p = \overline{-a, a}, s = \overline{-b, b}, \quad (3)$$

where  $(x(m), y(m))$  are the coordinates of the point  $(i, j)$  of the image  $\mathbf{Z}^s$  in the image  $\mathbf{Z}^d$  for the estimate  $\hat{h}_{i,j}^m$  and  $(2a+1) \times (2b+1)$  is the window size for calculation of CC. The optimal value  $\hat{h}_{(i,j)y}$  is determined similarly.

Joint processing of estimates allows compensating the inertia of the recursive estimation. The results given below were obtained with  $a = b = 1$ .

The joint processing of estimates  $\hat{\rho}_{i,j}^l, \hat{\rho}_{i,j}^r$  and  $\hat{\varphi}_{i,j}^l, \hat{\varphi}_{i,j}^r$  as well as finding the optimal values of  $\hat{\rho}_{i,j}$  and  $\hat{\varphi}_{i,j}$  are performed as described above using the same criteria of gradient estimation minimum and CC maximum.

Thus for joint processing of the estimates, depending on the criterion, we obtain two versions of the estimation algorithm:

*Algorithm A:* reverse processing using the criterion Eq. (2);

*Algorithm B:* reverse processing using the criterion Eq. (3).

## 2.2 Comparison of the estimation algorithm efficiency

Let us first consider the efficiency of the criterion Eq. (2). **Figure 2** shows an example of images used for the study. These are two adjacent frames of a video sequence where the vehicle located in the center is moving and the vehicle on the right is motionless. The parameters of inter-frame spatial shift of the moving vehicle are  $h_x = 3, h_y = 2, 95$  or  $\rho = 4, 2$ , and  $\varphi = 45^\circ$ .





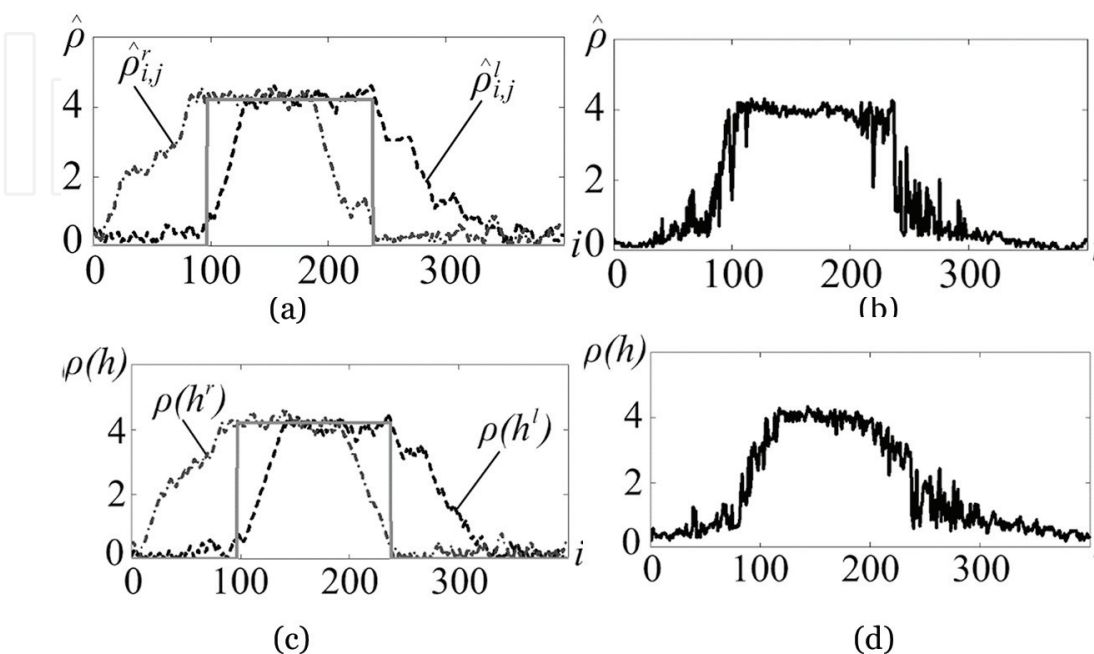
**Figure 2.**  
An example of adjacent frames of a video sequence with a moving object.

For example, **Figure 3** shows typical results of shift magnitude estimation for a single row of the reference image (**Figure 2**) while using Algorithm A. For a correct comparison of the estimates  $\hat{\rho}_{i,j}$ ,  $\hat{\varphi}_{i,j}$  and  $\hat{h}_{(i,j)x}$ , and  $\hat{h}_{(i,j)y}$ , the latter is recalculated to polar parameters:

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

**Figure 3a** shows dependencies  $\hat{\rho}_{i,j}^l$  and  $\hat{\rho}_{i,j}^r$  on  $i$  point in a row, **Figure 3b** shows the result of their joint processing, **Figure 3c** shows dependencies  $\rho(h^l)$  and  $\rho(h^r)$  on  $i$ , and **Figure 3d** shows the result of their joint processing. The solid gray line represents the true value of the deformation parameter. The results for parameters  $(\rho, \varphi)$  are visually better. Estimates given in **Table 1** also confirm that. **Table 1** shows mean value  $m_{\bar{h}}$  and variance  $\hat{\sigma}_{\bar{h}}^2$  of estimation error for the area with and without motion for results shown in **Figure 3b** and **d**.

Note that with the use of parameters  $(\rho, \varphi)$  in the motion area, the mean value of the estimation error is lower, and the variance is more than two times less than with the use of the set of parameters  $(h_x, h_y)$ . In the area without motion, the use of parameters  $(\rho, \varphi)$  is also preferable since the bias of the estimates is less despite the



**Figure 3.**  
An example of estimates of shift magnitude for a row using criterion Eq. (2).

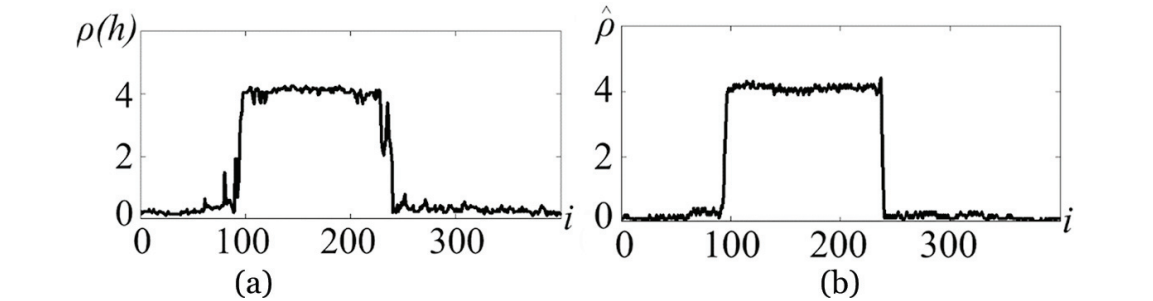
| Algorithm            | Motion area   |                   |   |                   | Area without motion |                   |   |                   |
|----------------------|---------------|-------------------|---|-------------------|---------------------|-------------------|---|-------------------|
|                      | $m_{\vec{h}}$ |                   | $\hat{\sigma}_{\vec{h}}^2 \times 10^{-2}$ |                   | $m_{\vec{h}}$       |                   | $\hat{\sigma}_{\vec{h}}^2 \times 10^{-2}$ |                   |
|                      | $(h_x, h_y)$  | $(\rho, \varphi)$ | $(h_x, h_y)$                              | $(\rho, \varphi)$ | $(h_x, h_y)$        | $(\rho, \varphi)$ | $(h_x, h_y)$                              | $(\rho, \varphi)$ |
| For a single row     |               |                   |   |                   |                     |                   |   |                   |
| A                    | 0.46          | 0.31              | 21.27                                     | 10.41             | 0.77                | 0.49              | 21.91                                     | 26.11             |
| B                    | 0.18          | 0.1               | 11.22                                     | 0.66              | 0.3                 | 0.15              | 3.82                                      | 0.89              |
| C                    | 0.37          | 0.06              | 19.8                                      | 4.84              | 0.44                | 0.22              | 3.4                                       | 1.42              |
| D                    | 0.06          | 0.01              | 0.38                                      | 0.26              | 0.03                | 0.02              | 0.29                                      | 0.03              |
| MVFAST               | 0.05          |                   | 25.3                                      |                   | 0.01                |                   | 0.01                                      |                   |
| For the entire image |               |                   |   |                   |                     |                   |   |                   |
| A                    | 0.61          | 0.53              | 23.6                                      | 14.3              | 0.73                | 0.45              | 23.8                                      | 18.8              |
| B                    | 0.42          | 0.34              | 7.3                                       | 1.7               | 0.29                | 0.15              | 2.6                                       | 1.5               |
| C                    | 0.13          | 0.14              | 15.9                                      | 10.1              | 0.46                | 0.25              | 2.8                                       | 2.9               |
| D                    | 0.07          | 0.01              | 1.4                                       | 0.69              | 0.09                | 0.02              | 0.28                                      | 0.02              |
| MVFAST               | 0.08          |                   | 18.6                                      |                   | 0.02                |                   | 0.05                                      |                   |

**Table 1.**  
 Estimation error of shift vectors of deformation field.

slightly larger variance. For comparison, **Table 1** also shows the results for the mean value and estimation of the variance  $\hat{\sigma}_{\vec{h}}^2$  of estimation errors averaged over the entire image.

**Figure 4** shows the results of the joint processing of the sets of estimates  $(\hat{h}_{(ij)x}, \hat{h}_{(ij)y})$  and  $(\hat{\rho}_{i,j}^l, \hat{\rho}_{i,j}^r)$  using the criterion Eq. (3) of the formation of the resulting estimate (Algorithm B). The results are for the same row as the results in **Figure 3**. **Figure 4a** corresponds to the set of parameters  $(h_x, h_y)$ , and **Figure 4b** corresponds to  $(\rho, \varphi)$ . **Table 1** summarizes the numerical data of the estimation error for the row and averaging over the entire image.

The results show that the use of CC maximum as the criterion can significantly improve the results of the joint processing of the estimates. For both sets of parameters, the mean value of the error for the motion area does not exceed 0.2. The variance of the error is approximately halved for the set of parameters  $(h_x, h_y)$  and 15 times less for parameters  $(\rho, \varphi)$ . For the area without motion, the mean error is less by about 2.6 and 3.3 times. The variance of the estimates also decreases. However, when using CC, the computational complexity of the estimation procedure increases significantly.



**Figure 4.**  
 The resulting estimates of shift magnitude for a row using criterion Eq. (3).

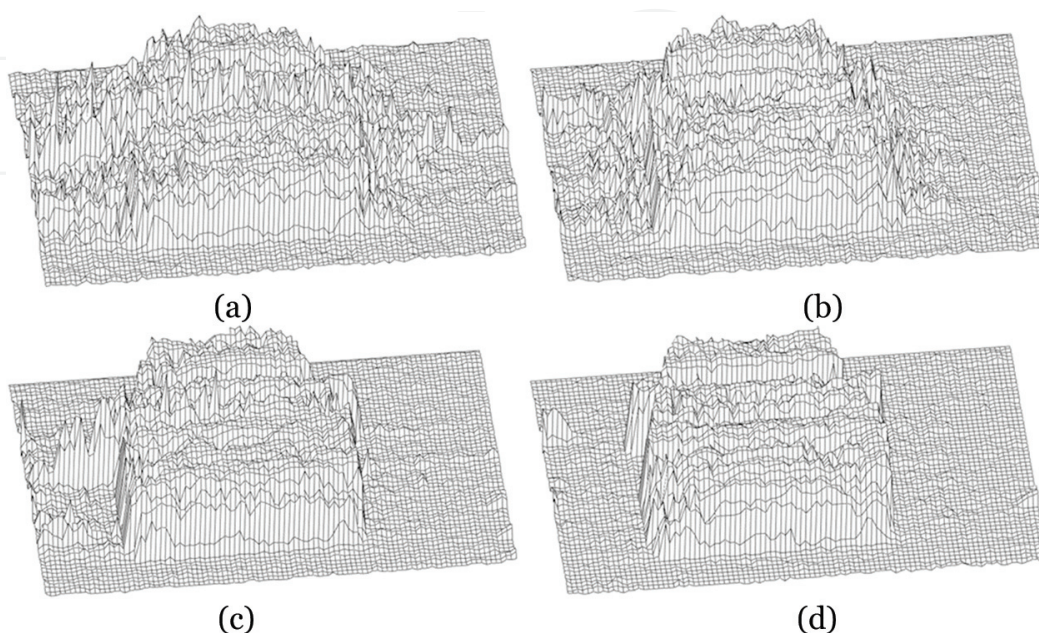
**Figure 5** shows the visualization of estimates of the shift vectors' field  $\mathbf{H}$  for the four estimation algorithms described above. It shows the magnitudes of the estimated vectors as a function of coordinates of nodes of the reference image. **Figure 5a** and **b** shows the results for criteria Eqs. (2) and (3) for the set of parameters  $(h_x, h_y)$  and **Figure 5c** and **d** for the same criteria when parameters  $(\rho, \varphi)$  are used. The figures illustrate and confirm the conclusions made above.

### 2.3 Use of inter-row correlation of adjacent rows of the image

The approach to estimate the field  $\mathbf{H}$  considered above processes images row by row as one-dimensional signals. Let us consider the possibility to increase processing efficiency by taking into account correlation of the adjacent rows of image [13, 17]. For each node  $(i, j)$  of the reference image, we will also form two estimates  $\hat{h}_{i,j}^1$  and  $\hat{h}_{i,j}^2$  of the parameters of the vector  $\bar{h}_{i,j}$ . To form the estimate  $\hat{h}_{i,j}^1$ , the stochastic gradient procedure processes rows one after the other with the change in direction after each row, for example, the odd rows from the left to the right, forming estimates  $\hat{h}_{i,j}^{1l}$ , and the even rows, from the right to the left, forming estimates  $\hat{h}_{i+1,j}^{1r}$ . The current shift estimate  $\hat{h}_{i+1,j}^1$  in a node is compared with the estimate  $\hat{h}_{i,j}^1$ , obtained from the previous row. If the estimate  $\hat{h}_{i,j}^1$  is considered better using some criterion, then it is taken as the current estimate.

As the criterion we use CC maximum Eq. (3) with the difference that here the set of possible values is limited by the estimates  $\hat{h}_{i+1,j}^1$  and  $\hat{h}_{i,j}^1$ . The estimate  $\hat{h}_{i,j}^2$  is also formed in a similar way with the difference that, when processing, the odd rows are processed from the right to the left and the even rows from the left to the right.

The joint processing of estimates  $\hat{h}_{i,j}^1$  and  $\hat{h}_{i,j}^2$  is carried out as described above. On the one hand, when taking into account inter-row correlation, the formation of  $\hat{h}_{i,j}^1$  and  $\hat{h}_{i,j}^2$  requires a greater amount of computation. On the other hand, the set of



**Figure 5.**  
 Visualization of estimating shift vectors' field by algorithms A and B with different criteria.



possible estimates for their joint processing turns out to be significantly smaller, which reduces the computational complexity.

Thus we have two more algorithms to estimate the shift vectors' field:

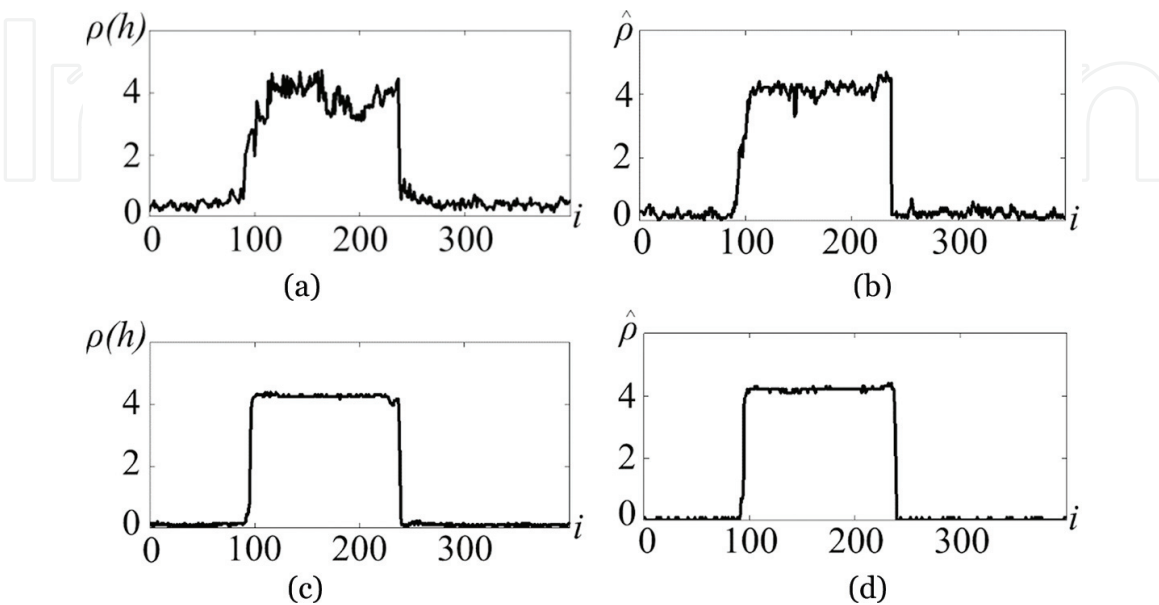
- Algorithm C: processing of adjacent rows using the criterion Eq. (2); and
- Algorithm D: processing of adjacent rows using the criterion Eq. (3).

## 2.4 Comparison of algorithms C and D

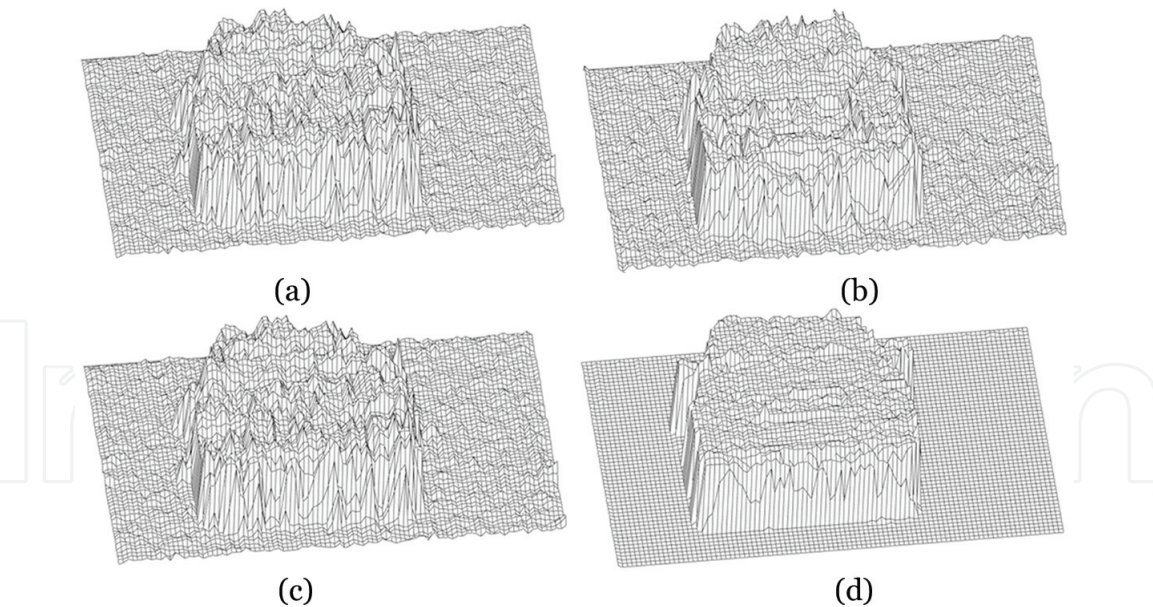
**Figure 6** shows the typical results of joint processing of the sets of estimates  $\hat{h}_{(ij)x}$ ,  $\hat{h}_{(ij)y}$  (**Figure 6a**) and  $\hat{\rho}_{i,j}$ , and  $\hat{\varphi}_{i,j}$  (**Figure 6b**) using Algorithm C. **Figure 6c** and **d** shows the results for the same sets and Algorithm D. It can be seen from the figures that the use of inter-row correlation can significantly improve the estimation results. For the parameters  $(h_x, h_y)$  and Algorithm C, mean value of the error for motion area decreases by 1.2 times, error variance by 1.1 times, and for Algorithm D by 3 times and 29 times, respectively. For the set of parameters  $(\rho, \varphi)$  and Algorithm C, the mean value of the error for motion area decreases by 5 times, variance by 2.1 times, and for Algorithm D by 10 times and 2.5 times, respectively. The numerical values are listed in **Table 1**.

**Figure 7** shows visualization of the estimates of shift vectors' field **H** for the estimation algorithms. **Figure 7a** and **b** corresponds to the algorithms C and D and set of parameters  $(h_x, h_y)$ , and **Figure 7c** and **d** corresponds to the same algorithms and the set of parameters  $(\rho, \varphi)$ . It can be seen that criterion Eq. (3) provides a greater accuracy of estimation (but also a larger computational costs). When maximum spatial accuracy is required, it is advisable to use one of algorithms B and D. The error for the Algorithm D is smaller, but the algorithm has a greater computational complexity.

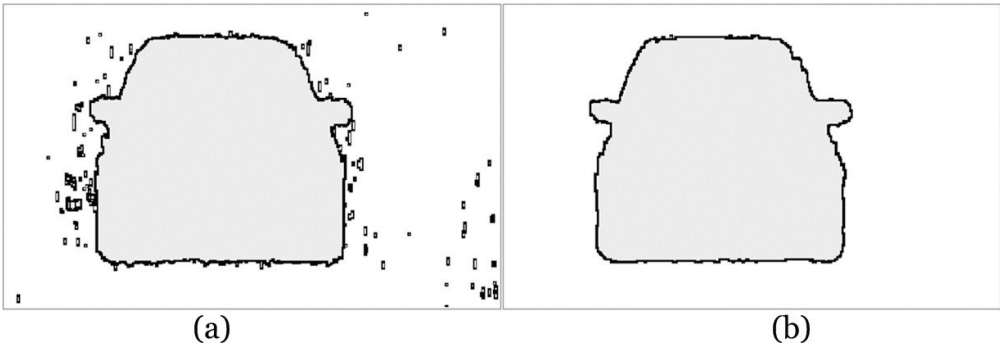
**Figure 8** shows moving object area and its contour obtained from the results of algorithms C (**Figure 8a**) and D (**Figure 8b**) by thresholding with the threshold equal to 0.1 of the maximum value of shift magnitude. Note that there are practically no errors of the second kind for Algorithm D.



**Figure 6.**  
Example of estimates of shift magnitude for a row using inter-row correlation.



**Figure 7.**  
 Visualization of shift vectors' field while using inter-row correlation.

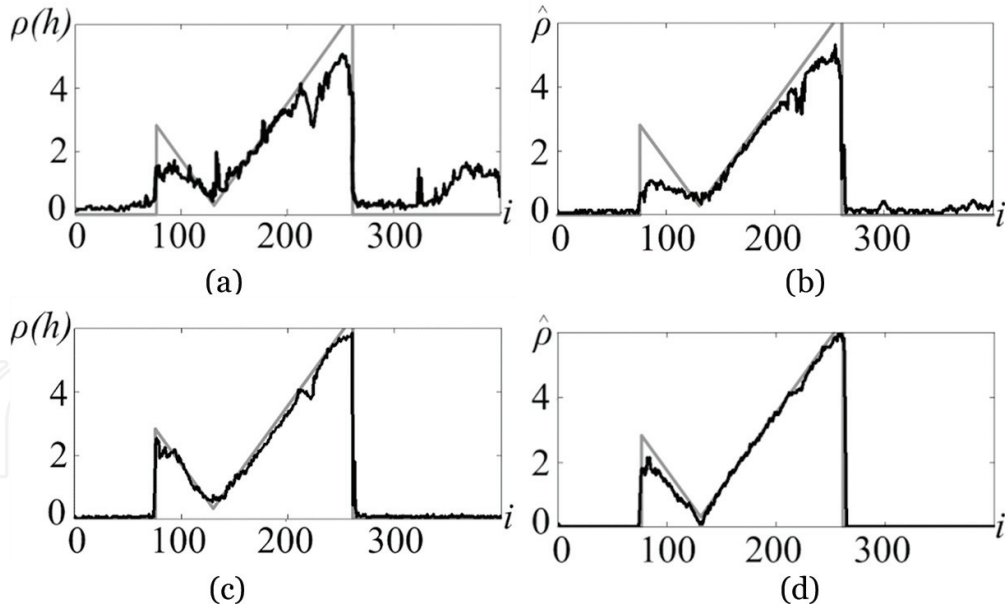


**Figure 8.**  
 Results of moving object area detection.

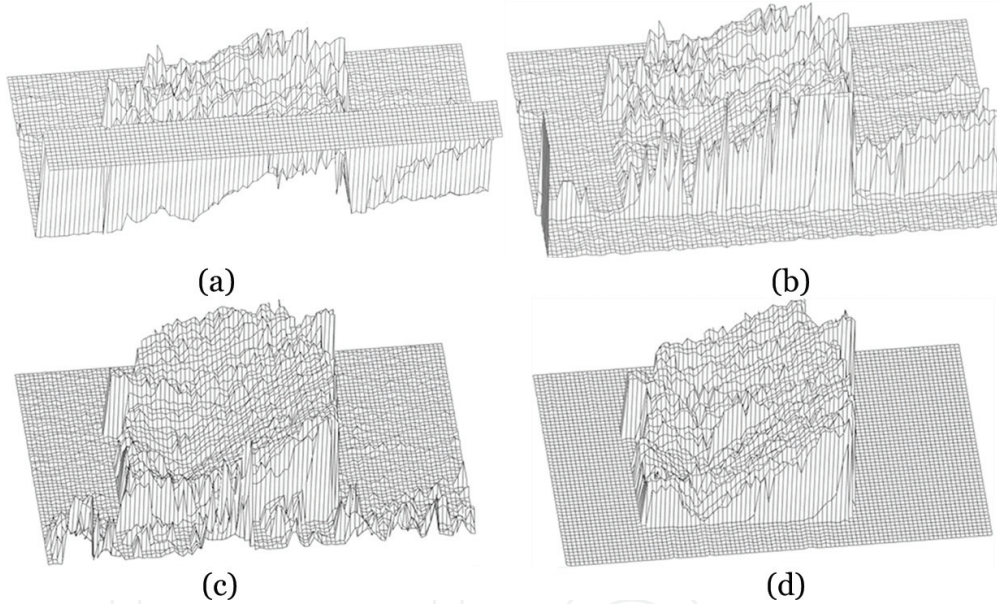
### 3. Efficiency of the algorithms for complex motion

Let us consider the possibilities of the developed algorithms when estimating shift vectors' field for more complex types of motion, in particular, for the case when inter-frame geometric deformations of images of a moving object can be described by a similarity model with a vector of parameters  $\bar{\alpha} = (\bar{h}, \varphi, \kappa)$ . To compare the results of estimation of the shift vectors' field with the results obtained for the images in **Figure 2**, the right image was modeled with the corresponding parameters of inter-frame deformations:  $\bar{h} = (2, 3)^T$ ,  $\varphi = 4^\circ$ , and  $\kappa = 1$ .

The results presented below are for algorithms B and D which were shown to have the highest efficiency with the use of Eq. (3) as a criterion. **Figure 9** shows typical results for a single row of the image; here graphs (a) and (b) correspond to the Algorithm B and the set of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$ , respectively, and graphs (c) and (d) correspond to Algorithm D. In this case the estimates for the set of parameters  $(\rho, \varphi)$  are closer to the true values (gray line). **Figure 10** shows visualization of estimates of the shift vectors' field for the entire image. **Figure 10a** and **b** corresponds to Algorithm B, and **Figure 10c** and **d** corresponds to Algorithm D with the set of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$ , respectively.



**Figure 9.**  
Example of estimates of shift magnitude for a row in case of complex motion of an object.



**Figure 10.**  
Visualization of shift vectors' field in case of complex motion of an object.

In the presented figures, in order to remove outliers significantly exceeding the true value of the shift, the estimation module was limited to  $\rho_{\max} = 10$ . When using Algorithm B, we note a significant number of false estimates  $\rho \geq \rho_{\max}$  both for and outside the area of motion. In particular, when using the parameters  $(h_x, h_y)$ , the estimates for a significant number of bottom rows of the image exceed the threshold. The use of the inter-row correlation significantly reduces the number of outliers for the set of parameters  $(h_x, h_y)$  and virtually eliminates them for the set  $(\rho, \varphi)$ .

#### 4. Comparison of the developed algorithms with the MVFAST algorithm

For block approach [18, 19], one of the most effective algorithms for the formation of shift vectors' field is the MVFAST (Motion Vector Field Adaptive Search



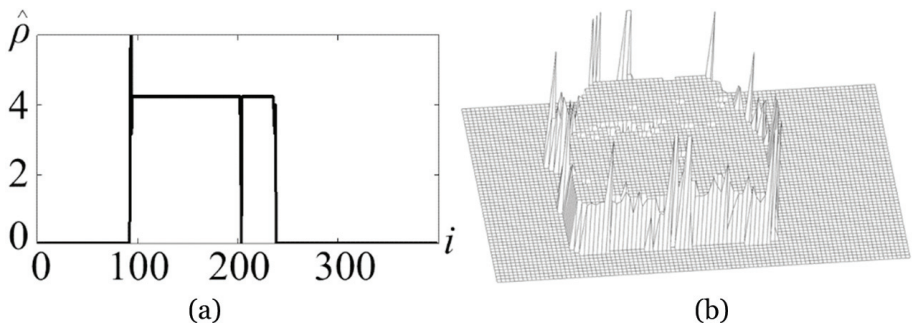
Technique) algorithm [20]. Therefore, we compare the effectiveness of the proposed algorithms and the MVFAST algorithm.

In the MVFAST algorithm, the deformed image is divided into many nonoverlapping square blocks of a predetermined size. In most cases the block sizes are  $2 \times 2$ ,  $4 \times 4$ , etc., but the algorithm allows setting blocks of any size, starting from one pixel. The estimation of the shift vector for the current block is determined by the estimates of the motion vectors of the neighboring blocks. The maximum value of the motion vector length of neighboring blocks is compared with two threshold values  $L_1$  and  $L_2$ , the values of which depend on the type of video sequence. As a result motion activity is classified as low, medium, and high. The criterion is the minimum of the sum of absolute differences (SAD). This characteristic for a block is the sum of the modules of the differences between the nodes of the current block and the corresponding nodes of the image on which the block location is searched. To improve performance and eliminate the influence of noise in the MVFAST algorithm, an early stop of the search is used. The search is terminated if the SAD is less than the specified threshold value  $T$ . Early stopping of the search can be disabled by setting the threshold  $T$  to zero.

Let us compare the results of estimation of the deformation field by algorithms D and MVFAST. For correctness of the comparison, below are the results obtained for the images in **Figure 2** with the parameters of Algorithm D,  $\lambda_h = \lambda_\rho = 0.1$ ,  $\lambda_\varphi = \pi/180$ ,  $\Delta x = \Delta y = 1$ ,  $\Delta_h = \Delta_\rho = 0.2$ , and  $\Delta_\varphi = 5\pi/180$ , and parameters of the MVFAST algorithm,  $L_1 = 1$ ,  $L_2 = 2$ ,  $T = 1$ , and block size  $1 \times 1$ . **Figure 11a** shows the estimates of the shift magnitude of the image points corresponding to the nodes of a single row of the reference image and formed by the MVFAST algorithm. The results of Algorithm D with the sets of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$  are shown in **Figure 6c** and **d**, respectively.

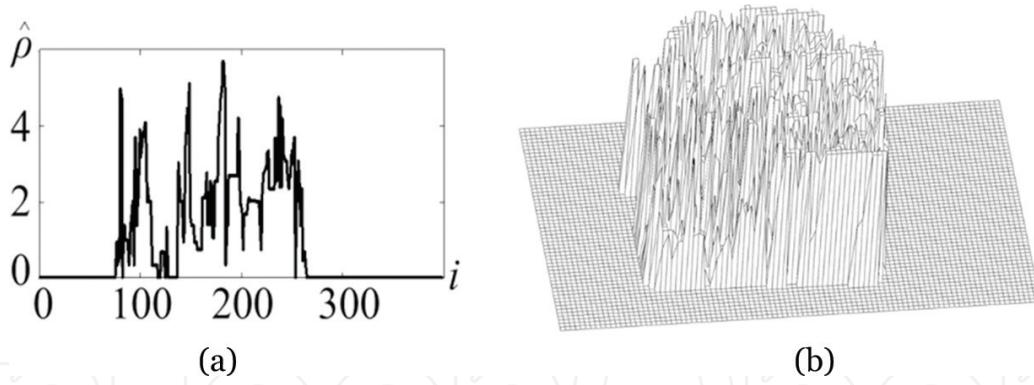
It can be seen from the figure that the MVFAST algorithm forms fairly accurate estimates of node shifts, but, unlike Algorithm D, it has errors at the boundaries of the image of an object. Errors also occur in low-contrast areas inside the object. The latter is due to the fact that SAD in the center of the search for them often does not exceed the threshold  $T$ . Setting the threshold  $T = 0$  when a block size is 1 pixel also does not solve this problem. Algorithm D, due to the inertia of changes of estimates, is deprived of this disadvantage.

The values of the mathematical expectation and variance of estimation errors when using the MVFAST algorithm for a row and for the entire image are shown in **Table 1**. The table shows that the mathematical expectation of the MVFAST algorithm for the motion area is almost the same as for Algorithm D with the set of parameters  $(h_x, h_y)$  but several times higher (about five times for a row, eight times for the entire image) than for Algorithm D with the set of parameters  $(\rho, \varphi)$ . The variance of the estimation error for the motion area for the MVFAST algorithm



**Figure 11.**  
Example of estimates for the shift magnitude for a single row and visualization of the deformation field generated by the MVFAST algorithm.





**Figure 12.** Example of estimates of shift magnitude for a single row and visualization of the deformation field formed by the MVFAST algorithm for complex motion.

significantly exceeds the variance for Algorithm D (13 times with the set of parameters  $(h_x, h_y)$  and 27 times with the set of parameters  $(\rho, \varphi)$ . In the absence of noise for area without motion, the MVFAST algorithm shows slightly better results compared to Algorithm D with the set of parameters  $(h_x, h_y)$ , but the results for Algorithm D with the set of parameters  $(\rho, \varphi)$  is better. This is well illustrated by **Figure 11b**, which shows the visualization of the deformation field estimates for the entire image. The results of using Algorithm D with the sets of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$  are shown in **Figure 7b** and **d**, respectively.

Let us also compare the efficiency of the algorithms with the complex motion for the same conditions and parameters of inter-frame deformations  $(\bar{h} = (2, 3)^T, \varphi = 4^\circ, \kappa = 1)$  which were used in the analysis of algorithms B and D.

**Figure 12a** shows the estimates of shift magnitudes for a single row of the reference image (the same row that was selected in the analysis of Algorithm D), when the MVFAST algorithm is used. The results of Algorithm D with the sets of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$  are shown in **Figure 9c** and **d**, respectively. The visualization of deformation field for the MVFAST algorithm is shown in **Figure 12b**. The deformation fields obtained by Algorithm D with the sets of parameters  $(h_x, h_y)$  and  $(\rho, \varphi)$  are shown in **Figure 10c** and **d**, respectively.

It can be seen from the figure that when estimating a deformation field for complex motion, the MVFAST algorithm gives significantly worse results in terms of spatial accuracy compared to the developed algorithms, and it is practically inapplicable for estimating the parameters of the motion trajectory.

Note that the increase of block size in the MVFAST algorithm allows to get rid of detection gaps inside the object image but still does not solve the problem of errors at its boundaries. At the same time, the increase of block size reduces the accuracy of moving object area detection. The proposed algorithms, in contrast to the MVFAST algorithm, provide subpixel accuracy while estimating the parameters of the shifts that is fundamentally impossible for the MVFAST algorithm.

## 5. Analysis of computational costs

Along with the accuracy of estimation, computational complexity of the algorithms is also important. In this case, the required amount of computations per pixel of the reference image with coordinates  $(i, j)$  consists of two components:

- Computational cost of finding estimates  $\hat{h}_{i,j}^l$  and  $\hat{h}_{i,j}^r$  or  $\hat{\rho}_{i,j}^l, \hat{\rho}_{i,j}^r$  and  $\hat{\phi}_{i,j}^l, \hat{\phi}_{i,j}^r$ ;
- Computational cost of obtaining the resulting estimate in the joint processing of estimates  $\hat{h}_{i,j}^l$  and  $\hat{h}_{i,j}^r$  or a set of estimates  $\hat{\rho}_{i,j}^l, \hat{\rho}_{i,j}^r$  and  $\hat{\phi}_{i,j}^l, \hat{\phi}_{i,j}^r$ .

The second component depends on many factors: complexity of the background, noise, number and speed of moving objects, etc. It can be evaluated with the given limitations of the particular problem being solved. Therefore, in the framework of this work, we restrict ourselves from analyzing the first component of computational costs with two sets of parameters for the estimated shift vectors  $(h_x, h_y)$  and  $(\rho, \varphi)$  for the three algorithms discussed above:

- Algorithm A: single row reverse processing
- Algorithm C: joint processing of adjacent rows using gradient estimation minimum criterion
- Algorithm D: joint processing of adjacent rows using CC maximum criterion

Detailed analysis of the computational cost requires taking into account not only algorithm itself but a large number of other factors (type of the computing device, time of memory access and for other operations, etc.). Many of these factors depend on the particular imaging device and computing resources. Therefore, we analyze only the computational complexity of the calculated ratios of the algorithms. We will consider the number of operations of addition (+), subtraction (−), multiplication (×), division (÷), tacking the root ( $\sqrt{\phantom{x}}$ ), trigonometric functions (sin, cos), and interpolations for a node ( $\tilde{z}$ ). **Table 2** shows the number of such operations needed to process one node of an image.

The total computational cost is defined as

$$N = c_{\pm}N_{\pm} + c_{\times}N_{\times} + c_{\div}N_{\div} + c_{\tilde{z}}N_{\tilde{z}} + c_{\sqrt{\phantom{x}}}N_{\sqrt{\phantom{x}}} + c_{\sin}N_{\sin}, \tag{4}$$

| Operation            | Number of operations |                   |              |                   |              |                   |
|----------------------|----------------------|-------------------|--------------|-------------------|--------------|-------------------|
|                      | Algorithm A          |                   | Algorithm C  |                   | Algorithm D  |                   |
|                      | $(h_x, h_y)$         | $(\rho, \varphi)$ | $(h_x, h_y)$ | $(\rho, \varphi)$ | $(h_x, h_y)$ | $(\rho, \varphi)$ |
| +, −                 | 12                   | 16                | 24           | 32                | 156          | 160               |
| ×                    | 6                    | 20                | 10           | 38                | 114          | 128               |
| ÷                    | —                    | —                 | —            | —                 | 4            | 4                 |
| $\tilde{z}$          | 8                    | 8                 | 16           | 16                | 44           | 44                |
| sin, cos             | —                    | 4                 | —            | 8                 | —            | 4                 |
| $\sqrt{\phantom{x}}$ | —                    | —                 | —            | —                 | 4            | 4                 |

**Table 2.**  
The number of operations needed to process a single node of an image.

where  $c_{\pm}, c_{\times}, c_{\div}, c_{\tilde{z}}, c_{\sqrt{\cdot}}, c_{\sin}$  are the coefficients characterizing the time needed to perform corresponding operations and  $N_{\pm}, N_{\times}, N_{\div}, N_{\tilde{z}}, N_{\sqrt{\cdot}}, N_{\sin}$  are the number of operations.

Equation (4) allows estimating the computational cost of the algorithms for particular computing facilities. For example, for a PC with an AMD Turion II X2 M500 processor:  $c_{\pm} = 1.4$ ,  $c_{\times} = 1.4$ ,  $c_{\div} = 13.6$ ,  $c_{\tilde{z}} = 63$ ,  $c_{\sqrt{\cdot}} = 135$ , and  $c_{\sin} = 212$  ns. Then for the set of parameters  $(h_x, h_y)$ , average time needed to process one node of an image is  $0.53 \mu\text{s}$  for Algorithm A,  $1.06 \mu\text{s}$  for Algorithm C, and  $3.74 \mu\text{s}$  for Algorithm D. The average time for parameters  $(\rho, \varphi)$  is  $1.4 \mu\text{s}$  for Algorithm A,  $2.8 \mu\text{s}$  for Algorithm C, and  $4.62 \mu\text{s}$  for Algorithm D.

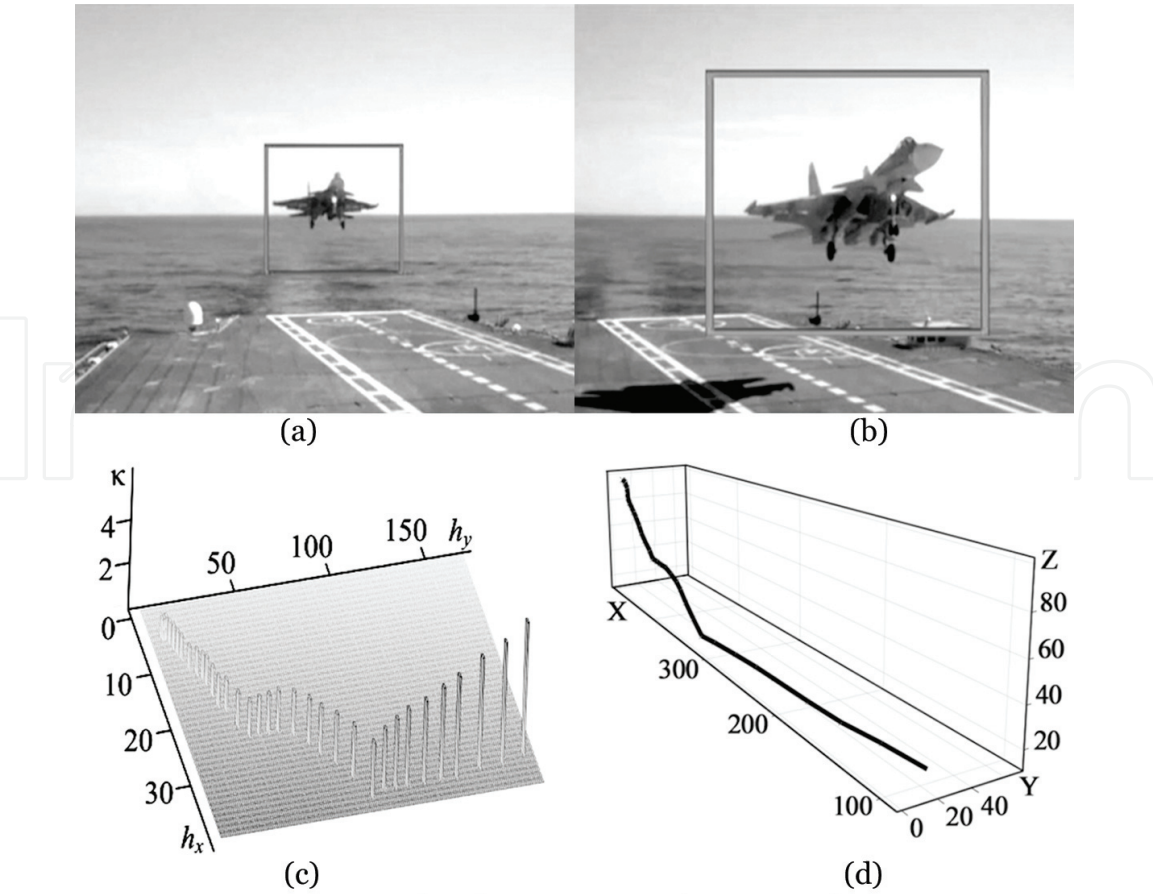
Let us note the features of the computational complexity of algorithms D and MVFAST. Studies have shown that the amount of computation of the latter is largely determined by the nature of the processed images. So if the adjacent frames of a video sequence do not contain motion, the computational complexity of the algorithm will be determined only by computing one SAD for each block. In the presence of motion, depending on its intensity, the computational complexity may increase in tens or even hundreds of times. In particular, when processing two frames of a video sequence without movement by the MVFAST algorithm with a block size of 1 pixel on the same PC, the average time spent on one node was  $0.64 \mu\text{s}$ . This is the lowest possible computational cost. The average processing time for a single node for the images shown in **Figure 2** was  $8.91 \mu\text{s}$ . For similar images with a complex type of motion, for example, with  $\bar{\alpha} = ((2, 3)^T, 4^\circ, 1)$ , processing time is  $21.84 \mu\text{s}$ . Note that the time expenses when using Algorithm D with the set of parameters  $(h_x, h_y)$  were  $7.44$ ,  $8.8$ , and  $10.75 \mu\text{s}$  under similar conditions, while with the parameters  $(\rho, \varphi)$ , the time expenses were  $8.13$ ,  $10.27$ , and  $11.8 \mu\text{s}$ .

From the above results, it can be seen that the computational complexity of Algorithm D weakly depends on the presence or absence of moving objects, their size, and type of motion. The computational complexity of the MVFAST algorithm is largely determined by these factors. So for the case when the inter-frame deformations of images of a moving object correspond only to a parallel shift, the computation time for algorithms D and MVFAST is approximately the same. When motion is complex, it increases in several times for the MVFAST algorithm.

## 6. An example of estimating the trajectory of a moving object using deformation field

Let us give an example of estimating the trajectory of a moving object using the developed algorithms. **Figure 13a** and **b** shows two frames from the video sequence being processed in which the aircraft lands on an aircraft carrier.

A complicating factor is the uneven movement of the camera toward the landing plane. Therefore, to estimate the position of the moving object relative to the scene (not to the camera), it was necessary not only to detect and identify the moving object area but also to stabilize the image. First, the distortion caused by the movement of the camera is compensated. To do this the inter-frame deformation of adjacent frames presented as similarity model  $\bar{\alpha} = (h_x, h_y, \varphi, \kappa)$  is estimated using deformation fields [21]. Then the area of a moving object is detected. Using the estimates of the deformation field in the area of moving object, the parameters of the inter-frame motion are found, which for 34 frames of the video sequence are isometrically presented in **Figure 13c**. These parameters are used to estimate a



**Figure 13.**  
*An example of estimation of moving object trajectory.*

three-dimensional trajectory of a moving object. **Figure 13d** shows the result in relative coordinates XYZ. The camera position at the initial moment of the shooting is taken as the origin.

## 7. Conclusion

The effectiveness of the motion area detection in sequence of images using pixel-by-pixel stochastic gradient estimation of the inter-frame shift vectors for all points of the reference image corresponding to its nodes (deformation field) is investigated. It is shown that stochastic gradient estimation of shift vectors does not give equivalent results for their representation as projections on the basic axes of the image and polar parameters. This is due to the fact that these sets of parameters have a different physical meaning. The analysis of variants of estimation algorithms for the deformation field has shown that the use of polar parameters provides higher accuracy of estimation.

Two approaches to estimate the parameters of the deformation field are considered. In the first approach, the stochastic gradient procedure sequentially processes all rows of an image to find estimates of shifts for all points of the reference image. It processes each row bidirectionally, i.e., from the left to the right and from the right to the left. The joint processing of the results allows compensating the inertia of the stochastic estimation. However, this approach does not take into account inter-row correlation, and images are processed as one-dimensional signals. In the second approach, to improve the accuracy of estimation, the correlation of image rows is taken into account. Algorithm processes rows one after the other with



change in direction after each row and uses obtained values to form resulting estimate for each node. From the results obtained for each node of the reference image, the resulting estimate of the shift vector is formed. Studies have shown that the second approach, with approximately the same computational complexity, showed significantly higher accuracy in the estimation of the shift vectors' field.

As a criterion for the formation of the resulting estimate, the minimum of gradient estimation of the objective function and the correlation maximum of local neighborhoods of the deformed and reference images were investigated. Correlation maximum criterion showed the best results. When using this criterion, not only the mean value and the variance of the estimation errors turned out to be less but also the effect of "oscillations" outside the motion area. But the use of this criterion requires a large computational cost.

The computational complexity of the proposed algorithms is investigated. Under equal conditions, the estimation of the projections of the shift vectors requires a slightly smaller amount of computations than the estimation of polar parameters. Using reverse processing of a row is less time-consuming compared with processing of the adjacent rows. The use of the correlation maximum criterion implies a larger amount of calculations than the use of the gradient estimation minimum criterion. The algorithms with large computational costs provide less error; therefore the choice of algorithm depends on the specific task.

The experiments have confirmed the effectiveness of using the developed algorithms to find the motion parameters and the trajectories of objects in a video sequence. In particular, an example of estimating three-dimensional trajectory of an aircraft in relative coordinates using 34 frames of a video sequence is considered. The parameters of the similarity model (parallel shift, rotation angle, and scale factor) were used to describe the inter-frame motion.

Comparison of the developed algorithms with the block algorithm MVFAST (Motion Vector Field Adaptive Search Technique) showed that under identical conditions, the latter shows worse results in accuracy of detecting the boundaries of motion. In addition, in contrast to the MVFAST algorithm, the proposed algorithms allow obtaining subpixel estimation accuracy.

Thus, the conducted study allows us to conclude that the pixel-by-pixel inter-frame estimation of the shift vectors of the reference image points corresponding to its nodes is a promising approach in detecting the area of a moving object in a series of images.

## **Acknowledgements**

The reported study was funded by the RFBR and Government of Ulyanovsk Region according to the research projects 18-41-730006 and 18-41-730011.

IntechOpen

### Author details

Alexander Tashlinskii<sup>1</sup> and Pavel Smirnov<sup>2\*</sup>

<sup>1</sup> Ulyanovsk State Technical University, Ulyanovsk, Russia

<sup>2</sup> Ulyanovsk Instrument Manufacturing Design Bureau, Ulyanovsk, Russia

\*Address all correspondence to: [rtcis@mail.ru](mailto:rtcis@mail.ru)

### IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Elhabian SY, El-Sayed KM, Ahmed SH. Moving object detection in spatial domain using background removal techniques. *Recent Patents on Computer Science*. 2008;1(1):32-54. DOI: 10.2174/1874479610801010032
- [2] Karasulu B, Korukoglu S. Moving object detection and tracking in videos. In: *Performance Evaluation Software*. Springer Briefs in Computer Science. USA, New York, NY : Springer, 2013. pp. 7-30. DOI: 10.1007/11612704\_16
- [3] Wang L, Yung NHC. Extraction of moving objects from their background based on multiple adaptive threshold and boundary evaluation. *IEEE Transactions on Intelligent Transportation Systems*. 2010;1(11): 40-51. DOI: 10.1109/TITS.2009.2026674
- [4] Kutsov RV, Trifonov AP. Detection of a moving object in the image. *Journal of Computer and Systems Sciences International*. 2006;3(45):459-468. DOI: 10.1134/S1064230706030129
- [5] Grishin SV, Vatolin DS, Lukin AS. The block methods review of motion estimation in digital video signals. *Programmnye Sistemy i Instrumenty*. 2008;1(9):50-62. [in Russian]
- [6] Zolotykh NY, Kustikova VD, Meerov IB. The review of searching and tracking methods in video. *Vestnik Nizhegorodskogo universiteta im N.I. Lobachevskogo*. 2012;5(2):348-358. [in Russian]
- [7] Tsyppkin YZ. *Information Theory of Identification*. Moscow: Fizmatlit; 1995. 336 p. [in Russian]
- [8] Tashlinskii AG. Computational expenditure reduction in pseudo-gradient image parameter estimation. *Lecture Notes in Computer Science*. 2003;2658:456-462. DOI: 10.1007/3-540-44862-4\_48
- [9] Tashlinskii AG. Pseudogradient estimation of digital images interframe geometrical deformations. In: Obinata G, Dutta A, editors. *Vision Systems: Segmentation & Pattern Recognition*. Austria, Vienna: I-Tech; 2007. pp. 465-494. DOI: 10.5772/4975
- [10] Tashlinskii AG. Optimization of goal function pseudogradient in the problem of interframe geometrical deformations estimation. In: Yin PY, editor. *Pattern Recognition Techniques, Technology and Applications*. Austria, Vienna: I-Tech; 2008. pp. 249-280. DOI: 10.5772/6244
- [11] Tashlinskii AG, Smirnov PV, Biktimirov LS. Methods of finding gradient estimates of target function for measurement of images parameters. *Pattern Recognition and Image Analysis*. 2011;2(21):339-342. DOI: 10.1134/S1054661811021057
- [12] Tashlinskii AG, Khoreva AM, Smirnov PV. Selection of finite differences in finding the stochastic gradient of the objective function in the estimation procedures of interframe deformations of images. *Radiotekhnika*. 2012;9:56-60. [in Russian]
- [13] Smirnov PV, Tashlinskii AG. Method for moving object area identification in image sequence. *Radiotekhnika*. 2015;6:5-11. [in Russian]
- [14] Tashlinskii AG, Kurbanaliev RM, Zhukov SS. Method for detecting instability and recovery of signal shape under intense noise. *Pattern Recognition and Image Analysis*. 2013; 3(23):425-428. DOI: 10.1134/S1054661813030139
- [15] Tashlinskii AG, Smirnov PV, Tsaryov MG. Pixel-by-pixel estimation of scene motion in video. *International Archives of the Photogrammetry, Remote Sensing and Spatial*

Information. 2017;**XLII-2/W4**:61-65.  
DOI: 10.5194/isprs-archives-XLII-2-W4-61-2017

[16] Tashlinskii AG, Smirnov PV. Pixel-by-pixel estimation of interframe geometric deformations of images in case selecting the area of a moving object. *Avtomotizatsiia Protsessov Upravleniia*. 2015;**1**(39):41-49. [in Russian]

[17] Tashlinskii AG, Smirnov PV. Moving object area identification in image sequence. In: *International Siberian Conference on Control and Communications (SIBCON)*; 21–23 May 2015; Russia. Omsk: IEEE; 2015. pp. 1-5. DOI: 10.1109/SIBCON.2015.7147239

[18] Bogush RP, Lysenko VI, Samoschenkov GA. Blocks-based algorithms combination for optical flow calculation for moving objects detection and tracking in video sequences. *Vestnik Polotskogo Gosudarstvennogo Universiteta. Fundamentalnye Nauki*. 2011;**4**:2-7. [in Russian]

[19] Grishin SV, Vatolin DS, Lukin AS. Review of block methods of motion estimation in digital video signals. *Programmnye Sistemy i Instrumenty*. 2008;**9**:50-62. [in Russian]

[20] Hosur PI, Ma KK. Motion vector field adaptive fast motion estimation. In: *Second International Conference on Information, Communications and Signal Processing (ICICS 1999)*; December 1999; Singapore. pp. 7-10

[21] Kaveev IN, Tashlinskii AG. Estimate parameters of the affine model of image binding by conjugate points. In: Negoda VN, editor. *Informatika, Modelirovanie, Avtomatizatsiia: Sbornik Nauchnykh Trudov*. Ulyanovsk: UlSTU; 2009. pp. 109-111. [in Russian]