Pseudogradient Estimation of Digital Images Interframe Geometrical Deformations

A.G. Tashlinskii
Ulyanovsk State Technical University
Russia

1. Introduction

Nowadays the systems of information extraction that include spatial apertures of signal sensors are widely used in robotics, for the remote exploration of Earth, in medicine, geology and in other fields. Such sensors generate dynamic arrays of data having the proper feature which is in their space-time correlation and due to which they can be represented in the form of multidimensional images (Gonzalez & Woods, 2002). When producing algorithmic software for the processing of such images it is necessary to take into account the dynamics of the scene to be observed, distortions caused by signal propagation environment, spatial movements of signal sensors and imperfection of their construction. The influence of the mentioned factors can be described through mathematical models of space-time deformations of multidimensional grids with the specified images.

The estimation of varying parameters of image spatial deformations is required not only in robotics applications, but also to solve a wide range of other problems, for example, for automated search of a fragment on the image, navigational tracking of mobile object course in the conditions of limited visibility, combination of multiregion images at remote explorations of Earth, in medical research. A lot of scientific calls for papers are devoted to different problems of image sequence space-time deformation parameters estimation (the bibliography is given, for example, in (Tashlinskii, 2000)). This chapter is devoted to one of the directions of solving this type of problems, where pseudogradient estimation of image interframe geometrical deformations (IIGDs) is considered.

Let us assume that the model of IIGDs is defined to an accuracy of a parameters vector $\overline{\alpha}$ and the estimation quality criterion is formulated in terms of some functional $J(\overline{\alpha})$ minimization showing expected losses. However, it is impossible to find optimal parameters $\overline{\alpha}^*$ in the mentioned sense in view of incompleteness of description of the images to be observed. In this case we can estimate the parameters $\overline{\alpha}$ on the basis of a given realization Z analysis of the image to be observed by means of some adaptation procedure which minimizes $J(\overline{\alpha}) = J(\overline{\alpha}, Z)$ for the given realization Z. However, it is reasonable to avoid this intermediate state of the research and estimate $\overline{\alpha}$ directly on values $J(\hat{\overline{\alpha}}, Z)$ (Polyak & Tsypkin, 1984):

$$\hat{\overline{\alpha}}_t = \hat{\overline{\alpha}}_{t-1} - \mathbf{\Lambda}_t \nabla J(\hat{\overline{\alpha}}_{t-1}, Z), \quad (1)$$

Source: Vision Systems: Segmentation and Pattern Recognition, ISBN 987-3-902613-05-9,
Edited by: Goro Obinata and Ashish Dutta, pp.546, I-Tech, Vienna, Austria, June 2007

where $\hat{\overline{\alpha}}_t$ – next after $\hat{\overline{\alpha}}_{t-1}$ approximation of the minimum point of $J(\hat{\overline{\alpha}}, Z)$; $\mathbf{\Lambda}_t$ – gain matrix (positively defined matrix determining a value of the estimates change at the t-th iteration); $\nabla J(\hat{\overline{\alpha}}_{t-1}, Z)$ – gradient of the functional $J(\hat{\overline{\alpha}}_{t-1}, Z)$. The necessity of multiple cumbersome calculations hinders the procedure (1) application in the image processing. It is possible to significantly reduce computational expenses due to the usage of contraction $\nabla J(\hat{\overline{\alpha}}_{t-1}, Z_t)$ instead of $\nabla J(\hat{\overline{\alpha}}_{t-1}, Z)$ at some part Z_t of realization Z at each iteration choosing, for example, Z_t in the form of a sliding window. For relatively large-sized images, the analysis of approaches (Tashlinskii, 2000; Minkina et al., 2007) to the synthesis of algorithms of IIGDs estimation in real time showed that it is expedient to seek a decision, satisfying the requirements of simplicity, rapid convergence and efficiency in various real situations, among recurrent non-identification algorithms. The pseudogradient algorithms (PGAs) constitute the most representative class of such algorithms.

The conception of the pseudogradient was introduced in work (Polyak & Tsypkin, 1973). A unified approach to the analysis and synthesis of various procedures of the stochastic minimization has been developed on the basis of it. For the given problem to be solved the pseudogradient $\overline{\beta}_t$ may be represented as any random vector in the parameter space depending on a function of losses and estimates $\hat{\overline{\alpha}}_{t-1}$ at the t-th iteration if the following condition is satisfied

$$\left[\nabla J(\hat{\overline{\alpha}}_{t-1}, Z)\right]^T M\{\overline{\beta}_t\} \geq 0, \tag{2}$$

where T – sign of transposition; $M\{\cdot\}$ – symbol of the mathematical expectation. In the geometrical interpretation the vector $\overline{\beta}_t$ is the pseudogradient if it makes, on average, an acute angle with the exact value of the functional $J(\hat{\overline{\alpha}}, Z)$ gradient. The class of PGAs includes algorithms of stochastic approximation, random search and many others.

The following procedure is used in PGA (Tsypkin, 1995)

$$\hat{\overline{\alpha}}_t = \hat{\overline{\alpha}}_{t-1} - \mathbf{\Lambda}_t \overline{\beta}_t, \tag{3}$$

where $\overline{\alpha}$ – vector of the parameters to be estimated; $t = \overline{1,T}$ – iteration number; $\hat{\overline{\alpha}}_0$ – initial approximation of the parameters vector; T – number of iterations. The algorithm is considered to be pseudogradient if $\overline{\beta}_t$ is the pseudogradient at each its iteration. In this case the iterations are, on average, performed in the direction of reduction of $J(\overline{\alpha})$ and sequence $\hat{\overline{\alpha}}_1, \hat{\overline{\alpha}}_2, \ldots, \hat{\overline{\alpha}}_t, \ldots$ converges to the optimal parameters when satisfying relatively weak conditions (Polyak, 1976).

If realizations α_t, $t = 1, 2, \ldots$, of the parameter α to be estimated are observed as, for example, in the problems of image brightness prediction, then we can choose $\beta_t = \hat{\alpha}_t - \alpha_t$ as the pseudogradient, where the estimate $\hat{\alpha}_t$ is found on realization Z or on a part of realization Z_t. In problems of image processing the quality functional $J(\overline{\alpha}, Z)$ is often expressed through the mathematical expectation of some function $f(\overline{\alpha}, Z)$:

$$J(\overline{\alpha}, Z) = M\{(f(\overline{\alpha}, Z))\}. \quad (4)$$

In particular, it can be mean square of error of some value χ:

$$f(\overline{\alpha}, Z) = (\hat{\chi}(\overline{\alpha}, Z) - \chi)^2 = \Delta^2(\overline{\alpha}, Z),$$

where χ – the exact value and $\hat{\chi}(\overline{\alpha}, Z)$ – its estimate. In this case the condition of the pseudogradientness is met if differentiation under the symbol of the mathematical expectation in (4) is possible.

We should also mention that the procedure (3) does not require compulsory finding $J(\hat{\overline{\alpha}}_{t-1}, Z_t)$ or $\nabla J(\hat{\overline{\alpha}}_{t-1}, Z_t)$, i.e. $J(\overline{\alpha})$ can be non-observable. It is necessary to meet only the condition of the pseudogradientness. At the non-observable realization an auxiliary observable quality functional $Q(\overline{\alpha})$ can be introduced and a noisy value $Q(\overline{\alpha})$ can be chosen as $\overline{\beta}_t$, whose point of extremum (not necessary the point of minimum) is obtained at the same optimal parameters. Later on, this chosen functional characterizing the estimation quality will be called the goal function. For example, when estimating the mathematical expectation of random value X the following value can be selected as the goal function

$$Q(\alpha, X) = M\{(X - \alpha)^2)\},$$

then, in the simplest case $\beta_t = x_t - \hat{\alpha}_{t-1}$, where x_t – value X at the t-th iteration. When estimating the correlation coefficient between the centered values X and Y the goal function can be represented as

$$Q(\alpha, X, Y) = M\{(\alpha X - Y)^2\},$$

then, $\beta_t = (\hat{\alpha}_{t-1} x_t - y_t)$, where x_t and y_t – realizations of X and Y at the t-th iteration.

The problem of IIGDs estimation considered in this chapter is related to the second type of problems, where it is necessary to use the auxiliary quality functional.

Let us note one more important property of the pseudogradient procedures that consists in that, $\overline{\beta}_t$ assumes dependence on estimation values $\hat{\overline{\alpha}}_p$, $p < t$ in the preceding samples and rows of the image that enables to run image processing in the order of some sweep. The last property is very important at practical realization of the algorithms.

Thus, to synthesize fast PGAs of parameters estimation $\overline{\alpha}$, it is necessary to find a relatively easily calculated pseudogradient of the given goal function of the estimation quality. In the next part, several possible ways of computational expense reduction when finding the goal function pseudogradient are considered.

2. Choice of pseudogradient

When synthesizing PGA the important stages are in the choice of a goal function and a rule of finding its pseudogradient. Let us consider some approaches to solve these problems.

Let the studied frames $\mathbf{Z}^{(1)} = \left\{z_{\bar{j}}^{(1)} : \bar{j} \in \Omega_{\bar{j}}\right\}$ and $\mathbf{Z}^{(2)} = \left\{z_{\bar{j}}^{(2)} : \bar{j} \in \Omega_{\bar{j}}\right\}$ of images specified on a regular samples grid $\Omega_{\bar{j}} = \left\{\bar{j} = (j_1, \ldots, j_n) : j_k = \overline{1, N_k}\right\}$ represent additive mixture of the informational pattern $\mathbf{X} = \left\{x_{\bar{j}}\right\}$ and a pattern $\Theta = \left\{\theta_{\bar{j}}\right\}$ of an independent noise:

$$\mathbf{Z}^{(1)} = \mathbf{X}^{(1)} + \mathbf{\Theta}^{(1)}, \quad \mathbf{Z}^{(2)} = \mathbf{X}^{(2)} + \mathbf{\Theta}^{(2)}, \tag{5}$$

where $\overline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T$ – vector of unknown geometrical transformation parameters of the image $\mathbf{X}^{(1)} = \mathbf{X}$ into the image $\mathbf{X}^{(2)} = \mathbf{X}\left(\bar{j}, \overline{\alpha}\right)$, for example, rotation, shift in some direction, scale change etc. In doing so, $x_{\bar{j}}$, $\theta_{\bar{j}}^{(1)}$ and $\theta_{\bar{j}}^{(2)}$ are homogeneous and have Gaussian distribution with zero mean and known covariance functions $R_{x\bar{j}\bar{l}} = M\left\{x_{\bar{j}} x_{\bar{l}}\right\}$;

$R_{\theta\bar{j}\bar{l}} = \sigma_\theta^2 \delta_{\bar{j},\bar{l}}$, where $\delta_{\bar{j},\bar{l}} = \begin{cases} 1, & if \quad \bar{j} = \bar{l}; \\ 0, & if \quad \bar{j} \neq \bar{l}. \end{cases}$ – Kronecker symbol; $\bar{j}, \bar{l} \in \Omega$.

Under the assumed constraints the goal function for the gradient estimation of the parameters vector $\overline{\alpha}$ can be written using the conditions of the optimal estimation obtained by means of the maximum likelihood method in work (Vasiliev & Tashlinskii, 1998). In particular, it is shown that if the image $\mathbf{Z}^{(1)}$ is noisy, then the maximization of the likelihood function is almost the same as minimization of the quadratic form. Then, for the gradient of the goal function we obtain

$$\nabla \mathrm{J}(\overline{\alpha}, Z) = \nabla\left[\left(\dot{\mathbf{Z}}_{\bar{j}}^{(2)} - \dot{\hat{\mathbf{X}}}\left(\bar{j}, \overline{\alpha}\right)\right)^T \dot{\mathbf{V}}_z^{-1}\left(\dot{\mathbf{Z}}_{\bar{l}}^{(2)} - \dot{\hat{\mathbf{X}}}\left(\bar{l}, \overline{\alpha}\right)\right)\right], \tag{6}$$

where $\mathbf{V}_z$ – covariance matrix of the conditional distribution $w\left(\left\{z_{\bar{j}}^{(2)}\right\} \middle| \mathbf{Z}^{(1)}, \overline{\alpha}\right)$, $\hat{\mathbf{X}}\left(\bar{j}, \overline{\alpha}\right)$ – prediction found on the basis of observations $\mathbf{Z}^{(1)}$, which is the best estimate in the sense of estimation error variance minimum of a deformed image . The point above the matrices denotes their lexicographic representation. In the same work it is shown that in many cases the product $\dot{\hat{\mathbf{X}}}^T(\overline{\alpha}) \dot{\mathbf{V}}_z^{-1} \dot{\hat{\mathbf{X}}}(\overline{\alpha})$ can be considered to be independent of the deformation parameters $\overline{\alpha}$. Then the gradient of the goal function is determined by the relation

$$\nabla \mathrm{J}(\overline{\alpha}, Z) = -\nabla\left[\dot{\hat{\mathbf{X}}}(\overline{\alpha}) \dot{\mathbf{V}}_z^{-1} \dot{\mathbf{Z}}^{(2)}\right]. \tag{7}$$

We should note that in the last case to find the optimal estimates of the parameters $\overline{\alpha}^*$ the maximization of the goal function is carried out. It requires performing of recurrent algorithm iterations not in the direction of the antigradient, but in the direction of the gradient which corresponds to minus in (7).

It is obvious, the expressions (6) and (7) can not be realized in systems of continuous image processing, because it requires great computational expenses. However, their simplification enables to obtain various realizable pseudogradients of the goal function. Let us consider some possible ways of such simplification. If we assume that the image insignificantly varies

from frame to frame (i.e. $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ are noisy realizations of the images $\mathbf{X}$ and $\mathbf{X}\left(\bar{j},\overline{\alpha}\right)$), then there is no need to calculate the unwieldy covariance matrix $\mathbf{V}_z$ of the conditional distribution $w\left(\left\{z_{\bar{j}}^{(2)}\right\}\middle|\,\mathbf{Z}^{(1)},\overline{\alpha}\right)$, because in this case $\mathbf{V}_z \approx \sigma_\theta^2 \delta_{\bar{j},\bar{l}}$, where σ_θ^2 – variance of additive noise according to the model of observations (5). However, in this case calculation of the optimal prediction $\hat{x}\left(\bar{j},\overline{\alpha}\right)$ requires matrix operations, which lead to very large computational expenses for large-sized images. We can obtain their reduction by substituting the optimal prediction of values of deformed frame for prediction at limited local region of image. We can achieve even more calculations reduction using various interpolations for the prediction. When performing the interpolations at the current iteration of the algorithm the estimates $\hat{\overline{\alpha}}$ obtained at the preceding iteration are employed (Minkina et al., 2007). Then, to find the pseudogradient at the t-th iteration of the algorithm it is enough to use a local sample $Z_t = \left\{z_{\bar{j},t}^{(2)}, \tilde{z}_{\bar{j},t}^{(1)}\right\}$ of samples, where $z_{\bar{j},t}^{(2)}$ – samples of the deformed image $\mathbf{Z}^{(2)}$ contained in a local sample at the t-th iteration and $\tilde{z}_{\bar{j},t}^{(1)} = \tilde{z}^{(1)}\left(\bar{j}_t, \hat{\overline{\alpha}}_{t-1}\right)$ – brightness values from continuous image $\tilde{\mathrm{Z}}^{(1)}$ obtained from $\mathbf{Z}^{(1)}$ through the chosen interpolation; $\bar{j}_t \in \Omega_{\bar{j},t} \in \Omega_{\bar{j}}$ – sample coordinates $z_{\bar{j},t}^{(2)}$ ($\Omega_{\bar{j},t}$ – plan of a local sample). The number of samples $\left\{z_{\bar{j},t}^{(2)}\right\}$ in Z_t will be called the local sample size and denoted with μ.

Under these assumptions the pseudogradients obtained on the basis of relations (6) and (7) will become

$$\overline{\beta}_t = \sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}_{\bar{j},t}^{(1)}}{\partial \overline{\alpha}} \Delta_{\bar{j},t} \Bigg|_{\overline{\alpha}=\hat{\overline{\alpha}}_{t-1}}, \tag{8}$$

$$\overline{\beta}_t = -\sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}_{\bar{j},t}^{(1)}}{\partial \overline{\alpha}} z_{\bar{j}t}^{(2)} \Bigg|_{\overline{\alpha}=\hat{\overline{\alpha}}_{t-1}}, \tag{9}$$

where $\Delta_{\bar{j},t} = \tilde{z}_{\bar{j},t}^{(1)} - z_{\bar{j},t}^{(2)}$.

We should note that the pseudogradient (8) is used for solving the problem of interframe difference mean square minimization. In this case it will be the goal function

$$\mathrm{J}\left(\overline{\alpha}, \mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\right) = \frac{1}{M} \sum_{\bar{j} \in \Omega} \left(z_{\bar{j}}^{(2)} - \tilde{z}_{\bar{j}}^{(1)}\right)^2,$$

where M – number of samples in the frame $\mathbf{Z}^{(2)}$.

The pseudogradient (9) corresponds to the problem of interframe correlation sample coefficient maximization:

$$\mathrm{J}\left(\overline{\alpha}, \mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\right) = \frac{\sum_{\bar{j} \in \Omega} \left(z_{\bar{j}}^{(2)} - z_m^{(2)}\right)\left(\tilde{z}_{\bar{j}}^{(1)} - \tilde{x}_m^{(1)}\right)}{M \hat{\sigma}_{z1} \hat{\sigma}_{z2}},$$

where: $\hat{\sigma}^2_{z1} = \frac{1}{M-1}\sum_{\bar{j}\in\Omega}\left(z^{(2)}_{\bar{j}} - z^{(2)}_{m}\right)^2$ and $\hat{\sigma}^2_{z2} = \frac{1}{M-1}\sum_{\bar{j}\in\Omega}\left(\tilde{z}^{(1)}_{\bar{j}} - \tilde{x}^{(1)}_{m}\right)^2$ – variance estimates of the images $\mathbf{Z}^{(2)}$ and $\tilde{\mathbf{Z}}^{(1)}$; $z^{(2)}_{\bar{j}m} = \frac{1}{M}\sum_{\bar{j}\in\Omega} z^{(2)}_{\bar{j}}$; $\hat{x}^{(1)}_{m} = \frac{1}{M}\sum_{\bar{j}\in\Omega}\tilde{z}^{(1)}_{\bar{j}}$; $\bar{j}\in\Omega_{\bar{j}}$.

Thus, in practical problems of IIGDs estimation the basic goal functions can be the interframe difference mean square and the interframe correlation sample coefficient. We should note that the pseudogradient (9) in contrast to (8) is invariant to the total variability of samples brightness of the image $\mathbf{Z}^{(2)}$. The choice of interframe difference mean square as the goal function is expedient in absence of multiplicative distortions and noncentered interference in the observable image models.

The vector $\overline{\beta}_t = \overline{\phi}(\nabla Q(\overline{\alpha}_{t-1}, Z_t))$ can be chosen as a pseudogradient, where $\overline{\phi}(\cdot)$ – vector function of the same dimensionality as ∇Q. For example, the function $\overline{\phi}(\cdot)$ can be linear transformation with the positively determined matrix. At that, if errors with respect to the true gradient have symmetric distributions with reference to zero, then the condition of the pseudogradientness (2) holds for any odd function $\overline{\phi}(\cdot)$. In particular, very simple and at the same time well converging algorithms of the parameters estimation are obtained when choosing the following sign function as $\overline{\phi}(\cdot)$ (Korn & Korn, 1968)

$$\overline{\beta}_t = \operatorname{sgn}\left(\nabla Q(\hat{\overline{\alpha}}_{t-1}, Z_t)\right), \tag{10}$$

where $\operatorname{sgn}(\nabla Q) = (\operatorname{sgn}(\nabla Q_1), \ldots, \operatorname{sgn}(\nabla Q_m))^T$. When using the peseudogradient (10) and the diagonal gain matrix in PGA (3) the i-th component of the vector $\hat{\overline{\alpha}}_t$ is different from the corresponding component of the vector $\hat{\overline{\alpha}}_{t-1}$ by $\pm\lambda_{i,t}$, where $\pm\lambda_{i,t}$ – corresponding diagonal element of the gain matrix $\mathbf{\Lambda}_t$; $i = \overline{1, m}$. At that, PGA iterations are carried out at finite and a priori known number of directions of the space of the parameters to be estimated. If each component of the error (10) in relation to the true gradient takes positive and negative values with equal probabilities, then the pseudogradientness condition (2) is met. Let us note the algorithms that use the pseudogradients of type (10) have wide application in various problems requiring IIGDs estimation in the conditions of complex noise assemblage.

3. Pseofogradient algorithms for interframe geometrical deformations parameters estimation

3.1 Algorithms at given set of geometrical deformations model parameters

If a parameters set $\overline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T$ of possible IIGDs is known, then at chosen goal function the problem amounts to estimation of their values that are constant on the images $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$. For example, if for (3) the interframe difference mean square is chosen as a goal function and its pseudogradient is given by relations (8) and (10), then to estimate $\overline{\alpha}$ we accordingly obtain the following algorithms:

$$\hat{\bar{\alpha}}_t = \hat{\bar{\alpha}}_{t-1} - \mathbf{\Lambda}_t \left(\sum_{\bar{j} \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}^{(1)}_{\bar{j},t}}{\partial \bar{\alpha}} \Delta_{\bar{j},t} \right)_{\bar{\alpha}=\hat{\bar{\alpha}}_{t-1}} ; \quad (11)$$

$$\hat{\bar{\alpha}}_t = \hat{\bar{\alpha}}_{t-1} - \mathbf{\Lambda}_t \operatorname{sgn}\left(\sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}^{(1)}_{\bar{j},t}}{\partial \bar{\alpha}} \Delta_{\bar{j},t} \right)_{\bar{\alpha}=\hat{\bar{\alpha}}_{t-1}} . \quad (12)$$

Experimental study show it is expedient to extend the algorithms set (11)–(12) by adding two more ones

$$\hat{\bar{\alpha}}_t = \hat{\bar{\alpha}}_{t-1} - \mathbf{\Lambda}_t \left(\sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}^{(1)}_{\bar{j},t}}{\partial \bar{\alpha}} \operatorname{sgn} \Delta_{\bar{j},t} \right)_{\bar{\alpha}=\hat{\bar{\alpha}}_{t-1}} ; \quad (13)$$

$$\hat{\bar{\alpha}}_t = \hat{\bar{\alpha}}_{t-1} - \mathbf{\Lambda}_t \left(\sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \Delta_{\bar{j},t} \operatorname{sgn} \frac{\partial \tilde{z}^{(1)}_{\bar{j},t}}{\partial \bar{\alpha}} \right)_{\bar{\alpha}=\hat{\bar{\alpha}}_{t-1}} . \quad (14)$$

In the algorithm (11) all the components of estimation increment vector depend on interframe differences $\Delta_{\bar{j},t}$, $\bar{j}_t \in \Omega_{\bar{j},t} \in \Omega_{\bar{j}}$. It determines higher estimation convergence speed compared to other given algorithms. However, at finite number of iterations the precision of (11) is often lower because we can not always attain little $\Delta_{\bar{j},t}$ and then, the estimates variation steps can be too large. In the algorithm (13) only signs of $\Delta_{\bar{j},t}$ are used. It is preferable to apply it when we want to avoid excessive influence of modulo large values $\Delta_{\bar{j},t}$, for example, in presence of infrequent but intensive impulse interference on image. The algorithm (12) is even more immune to interference, but it may not operate well in the neighborhood of zero values $\Delta_{\bar{j},t}$. If high accuracy of estimation is attained at some iteration, then the next step can be taken in the direction backward from optimal values of the parameters. To eliminate this disadvantage we can employ a sign function with expanded zero:

$$\operatorname{sgn}_\varepsilon(x) = \begin{cases} -1, & x < -\varepsilon, \\ 0, & |x| \le \varepsilon, \\ 1, & x > \varepsilon, \end{cases} \quad \varepsilon > 0 .$$

In the algorithm (14) the increase of convergence speed at large values $\Delta_{\bar{j},t}$ combines with immunity to derivatives estimation errors. As a result this algorithm is more resistant to interference than the algorithm (11).

When choosing the interframe correlation sample coefficient as a goal function the properties of the corresponding algorithms are close to the properties of the algorithms (11)–(14). The advantage in this case is in high immunity to additive noise and close to linear brightness distortions. Among disadvantages are larger computational expenses (determined by large size of the local sample) and also high sensibility to local extremums of the goal function.

It is necessary to note that the convergence speed of the estimates, formed by PGAs, is higher if the sequence of local samples, which is the basis for parameters estimation, is not correlated. To reduce correlation of the observations sequence it is expedient to choose random coordinates of samples of the local sample.

The algorithms (11)–(14) demonstrated high efficiency when estimating interframe deformations of simulated and real images. In particular, for two-dimensional images of size 64x64 pixels formed by means of the wave model (Krasheninnikov, 2003) shifted by several steps of sample grid and turned at any angle, the shift was estimated with error variance of about $2 \cdot 10^{-4}$ steps of the sample grid and rotation – $5 \cdot 10^{-5}$ radians. Let us give the results of analytical calculation of the probability $P(\Delta)$ of parallel shift estimate $\hat{h}_1$ error spillover of the given interval $\Delta = [-a\ a]$. The calculation was carried out on the basis of the accuracy analysis method of PGAs estimates at finite number of iterations (Tashlinskii & Tikhonov, 2001) for the PGA (12) and the following parameters: the images with a Gaussian brightness distribution and the autocorrelation function with correlation radius equal to 5; the signal variance-to-noise variance ratio $g = 100$; local sample size $\mu = 10$; initial error of the shift is $\overline{h}_o = (5, 4)^T$; $a = 1.0$, 0.3 и 0.1 (here, by correlation radius we imply the distance in steps of the sample grid when the autocorrelation function of the image is equal to 0.5). The value of estimate shift increment in one case was chosen to be constant $\lambda_{1,t} = \lambda_{2,t} = \lambda = const = 0.1$ and in the other – falling off according to the rule $\lambda_{1,t} = \lambda_{2,t} = \mathrm{var} = \dfrac{1}{(10 + 0.01t)}$. The plots of the probability $P(\Delta)$ are shown in Fig. 1. It follows from the analysis of the plots that at constant λ the balance between tendency of the estimate to the true value and error, caused by λ, comes at a certain iteration. The further increase of iterations number does not lead to estimates improvement. It enables to find the minimum number of iterations that is necessary to achieve the highest possible accuracy of parameter estimation.

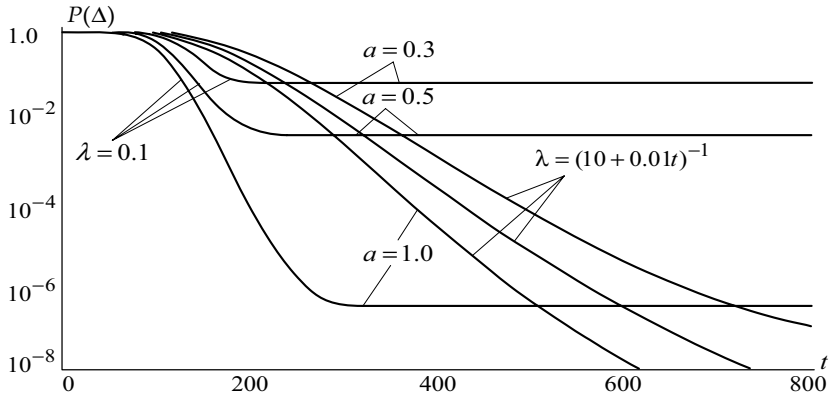


Fig. 1. Probability of estimate error spillover of the confidence interval

For the above-mentioned characteristics of images and PGA in Fig. 2 the shift h_1 estimate error probability distribution change at both constant (Fig. 2,a) and varying (Fig. 2,b) value

of the shift estimate increment is shown. For clearness the distributions are presented only for 12 iterations from the range of 10 to 100. The interframe difference mean square at local sample size $\mu = 4$ was used as a goal function. The analysis of the plots shows that at constant shift increment step the process of probability distributions forming stabilizes after about 500th iteration. At varying shift increment step the process of probability distribution forming does not have an equilibrium state and the estimate variance theoretically permanently decreases.

Let us note that at known set of IIGD parameters the algorithms (11)–(14) have shown a good performance at automated search of local fragments on images.

3.2 Algorithms at unknown set of geometrical deformations model parameters

If the form of IIGD is not given then we can specify a certain sample grid deformations model

$$\overline{\alpha}(\bar{j}) = \left(\bar{j} + \overline{h}_{\bar{j}}\right) = (j_1 + h_1, j_2 + h_2, \ldots j_n + h_n), \tag{15}$$

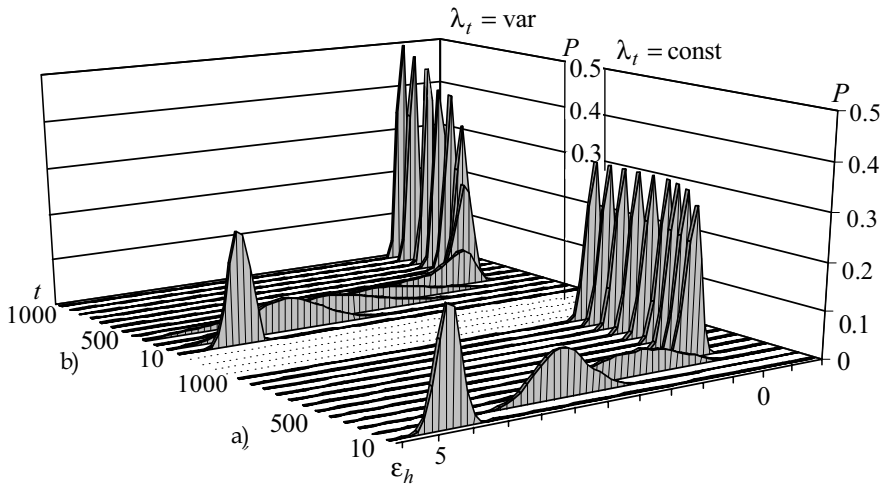


Fig. 2. Probability distributions of the shift estimate error at constant and varying estimate increment steps

considering its parameters to be varying, where $j_k \in \bar{j} \in \Omega_{\bar{j}}$, $j_k = \overline{1, N_k}$, $k = \overline{1, n}$; $\Omega_{\bar{j}}$ – n-dimensional rectangular grid. Then, the algorithm (3) can be written in the form

$$\hat{\overline{h}}_t = \hat{\overline{h}}_{t-1} - \mathbf{\Lambda}_t \overline{\beta}_t\left(Z_t, \bar{j}_t, \hat{\overline{h}}_{t-1}\right) \tag{16}$$

In this case to ensure variability of the estimates the components of the gain matrix $\mathbf{\Lambda}_t$ in (16) have to be bounded below. Then, assuming $\mathbf{\Lambda}_t = \mathbf{\Lambda}$ and choosing the interframe difference mean square as a goal function, we come to the algorithm

$$\hat{\bar{\alpha}}_t = \hat{\bar{\alpha}}_{t-1} - \mathbf{\Lambda}_t \left(\sum_{\bar{j}_t \in \Omega_{\bar{j},t}} \frac{\partial \tilde{z}^{(1)}(\bar{j}_t, \bar{h})}{\partial \bar{h}} \left(\tilde{z}^{(1)}(\bar{j}_t, \bar{h}) - z^{(2)}_{\bar{j},t} \right) \right)_{\bar{h}=\hat{\bar{h}}_{t-1}} . \quad (17)$$

In work (Tashlinskii, 2000) it is shown that if point-to-point correlation is within the limits $0.8 \div 0.99$ and signal variance-to-noise variance ratio is more than 50, then for many problems of IIGDs estimation it is enough to choose $\mu = 1$ in (17). In spite of the simplicity the algorithm (17) is rather effective at high correlation of the shifts $\bar{h}_{\bar{j}}$ in the direction of image scanning and relatively minor interframe brightness distortions. However, only one-dimensional filtering of deformation parameters that does not take into account interrow and interframe correlation is carried out in it. The simplest way to take into account this correlation can be in refinement of the estimates, obtained at one pass through the images. For that we can perform repeated passes at lesser values λ in the backward and other directions (along columns, diagonals) during which the obtained estimates are corrected.

The algorithms of the type (17) have shown a good performance at automatic combination of image fragments that have reciprocal spatial and amplitude deformations (the problem of image «pasting»). This problem often occurs when forming a unified image from a sequence of frames, obtained from a mobile object, that have small common regions on the adjacent frames. When solving the mentioned problem it is required as a rule to ensure continuity of spatial and brightness characteristics on the resulting image. The considered algorithms enable to do it. To illustrate it in Fig.3 an example of «pasting» of images is presented, where a) and b) – are the images of size 100×160 elements to be connected, having parallel shift $\bar{h} = (-0.5, -0.2)^T$, rotation angle $\varphi = 0.5^o$ and scale coefficient $k = 1.2$; c) – the result of «pasting» before spatial correction using the obtained estimates; d) – the result of «pasting» after correction. The estimates accounted for $\hat{\bar{h}} = (-0.676, -0.13)$, $\hat{\varphi} = 0.509^o$, $\hat{k} = 1.18$.

a) b) c) d)

Fig. 3. An example of automated «pasting» of images

Basically, the model (15) enables to define any IIGDs. However, if at the chosen order of the images pass the shifts $\bar{h}_{\bar{j}}$ change rapidly then their estimation by means of PGA (17) is difficult. It is due to the fact that when variation speed of the shifts increases in the direction of the image pass it is necessary to increase steps $(\mathbf{\Lambda}_t \bar{\beta}_t)$ of PGA (17). The last, in its turn, leads to estimation error increase. In this situation we can not improve the estimates even by repeated passes. The mentioned contradiction can be solved due to the usage at each following pass information about estimates, obtained at the preceding passes. Let us

consider the algorithm that forms the deformation matrix $\mathbf{H}_l$ of size $N_1 \times N_2 \times ... \times N_n$ (where n – image dimensionality) as an example of such an approach. This matrix contains shifts estimates $\hat{\bar{h}}_{\bar{j}}^{(l)}$ of all image pixels, corresponding to sample grid nodes of the frame $\mathbf{Z}^{(1)}$ after the l-th pass. The method of estimates forming can be various, for example, it can be determined by available conceptions regarding physical nature of geometrical deformations. Assuming, that all elements of the matrix $\mathbf{H}_{l-1}$ at the $(l-1)$-th pass have been determined we can write

$$\mathbf{H}_l = \left\| \hat{\bar{h}}_{\bar{j}}^{(l)} \right\| = f\left(\hat{\bar{h}}_{\bar{j}}^{(l-1)}, \left\{ \hat{\bar{\alpha}}_t^{(l)} \right\} \right),$$

$$\hat{\bar{\alpha}}_t^{(l)} = \hat{\bar{\alpha}}_{t-1}^{(l)} - \mathbf{\Lambda}_{lt} \bar{\beta}\left(z_{\bar{j},t}^{(2)}, \tilde{z}^{(1)}\left(\bar{j}_t + h_{\bar{j},t}^{(l-1)} \right), \hat{\bar{\alpha}}_{t-1}^{(l)} \right),$$

where $\hat{\bar{\alpha}}_t^{(l)} = \left(\hat{\alpha}_{1,t}^{(l)}, \hat{\alpha}_{2,t}^{(l)}, ..., \hat{\alpha}_{m,t}^{(l)} \right)^T$; $\bar{j}_t \in \Omega_t^{(l)} \in \Omega$; $\mathbf{\Lambda}_{l,t} = \left\| \begin{matrix} \lambda_{1,t}^{(l)} & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & \lambda_{n,t}^{(l)} \end{matrix} \right\|$; $z_{\bar{j},t}^{(2)}$, $\tilde{z}^{(1)}\left(\bar{j}_t + h_{\bar{j}t}^{(l-1)} \right) \in Z_t^{(l)}$; $Z_t^{(l)}$ – local sample of the goal function for the pseudogradient estimation at the t-th iteration of the l-th pass of the algorithm; $l = \overline{1,L}$ - pass number; L – given number of passes. Choosing various goal functions and pseudogradients we can obtain different algorithms. For example, if we choose the pseudogradient (10) for the interframe difference mean square and assume $\lambda_{it}^{(l)} = \lambda_l$; $\hat{\bar{h}}_{\bar{j}}^{(l)} = \hat{\bar{h}}_{\bar{j}}^{(l-1)} + \hat{\bar{\alpha}}_{\bar{j}}^{(l)}$, where $\hat{\bar{\alpha}}_{\bar{j}}^{(l)}$ – current estimate $\hat{\bar{\alpha}}^{(l)}$ at the point $\bar{j}$ then we obtain

$$\mathbf{H}_l = \left\| \hat{\bar{h}}_{\bar{j}}^{(l)} \right\|, \ \hat{\bar{h}}_{\bar{j}}^{(l)} = \hat{\bar{h}}_{\bar{j}}^{(l-1)} + \hat{\bar{\alpha}}_{\bar{j}}^{(l)}, \ l = \overline{1,L},$$

$$\hat{\bar{\alpha}}_t^{(l)} = \hat{\bar{\alpha}}_{t-1}^{(l)} - \lambda_l \operatorname{sgn}\left(\sum_{\bar{j}_t \in \Omega_{\bar{j},t}^{(l)}} \frac{\tilde{z}^{(1)}\left(\bar{j}_t + \hat{\bar{h}}_{\bar{j},t}^{(l-1)} + \hat{\bar{\alpha}}_{t-1}^{(l)} \right)}{\partial \bar{\alpha}} \left(\tilde{z}^{(1)}\left(\bar{j}_t + \hat{\bar{h}}_{\bar{j},t}^{(l-1)} + \hat{\bar{\alpha}}_{t-1}^{(l)} \right) - z_{\bar{j},t}^{(2)} \right) \Bigg|_{\bar{\alpha} = \hat{\bar{\alpha}}_{t-1}^{(l)}} \right). \quad (18)$$

If IIGDs with known parameters set (for example, the common for all image shift, rotation etc.) are present along with deformations of unknown type, then values of these parameters can be estimated, specified and taken into account when forming elements of matrix $\mathbf{H}_l$ at each algorithm pass.

Another advantage of the algorithms of the type (18) is that they enable to estimate IIGDs that do not satisfy the continuity requirement.

An example of such estimation is shown in Fig. 4, where a) and b) – images of size 256×256 elements having reciprocal shifts ($h_1 = 1.5$, $h_2 = 3.5$), besides in the lower image the continuity of geometrical deformations is violated (5 rows are missing) and the fragment is

absent; c) – result of parameter h_1 estimation at $L=40$ and $\mu=1$. The sudden change corresponding to the break of the parameter h_1 is well visible. Besides in the region, corresponding to the absent fragment, the estimates have significantly differing statistical properties and due to which these regions can be easily identified.

One of the disadvantages of PGA at IIGD parameters estimation is in a relatively minor definition domain of parameters, where effective convergence of estimates is ensured (not large operating range). The size of this region is determined by sample correlation that can appear in the local sample Z_t. The situation is also complicated by the fact that in real images samples of reference and deformed images taken rather far from each other are almost non-correlated. At operating with real images another serious disadvantage of PGAs is in the possibility of the estimates to converge to points of false extremums of the goal function in the parameter space.

3.3 Algorithms with adaptive forming of local sample size

In view of random character of images and noise, the estimate of the goal function is not unimodal and besides the global extremum it also contains false local extremums. The local extremums appear because of correlatedness of separate extensive objects on the image and are exposed if a portion of samples of the local sample appears into these regions, i.e. they are caused by limited size of the local sample. As local sample increases or changes the probability of this effect appearance sharply decreases. As a result it is reasonable to verify on the goal function local extremums attributes at each iteration of estimation and if any, to increase sample size or change it. Here, the sample size μ becomes an adaptive value.

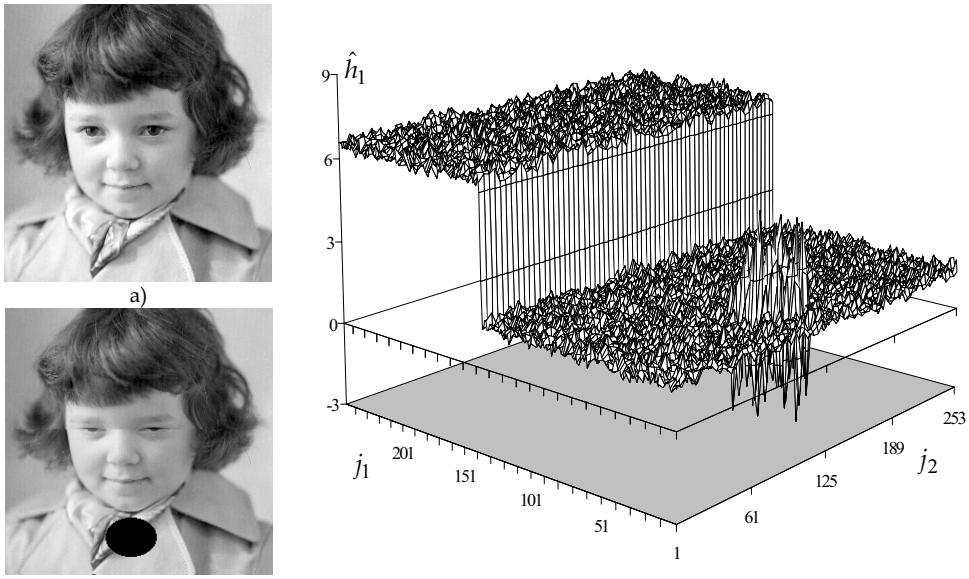


Fig. 4. An example of geometrical deformations estimation that does not satisfy the continuity requirement

Let us consider only one example of construction of IIGD parameter estimation PGA, where μ is adjusted automatically during the procedure performing at each iteration. A current iteration of parameter estimation is carried out when a certain condition is met. If at minimal μ for the current iteration the condition is not met, then μ increases step-by-step until the condition is fulfilled. So for the local sample, formed at the given iteration, its size is to be minimum to meet the condition of the iteration realization. For definiteness, we assume that the PGA (3) with the pseudogradient (10) and the diagonal gain matrix $\mathbf{\Lambda}_t$ is used. Then, the estimates for the i-th parameter are formed according to the following rule:

$$\hat{\alpha}_{i,t+1} = \hat{\alpha}_{i,t} \mp \lambda_{t+1} \operatorname{sgn}\left(\beta_{i,t+1}\left(\hat{\bar{\alpha}}_t, Z_t\right)\right), \quad t = \overline{1,T}, \quad i = \overline{1,m},$$

where the signs «–» and «+» correspond to the problems of minimization and maximization of the goal function.

Given a certain initial sample size $\mu_{t\min}$, whose minimum value at interframe correlation sample coefficient maximization must be not less than 2 and at the interframe difference mean square minimization – is equal to 1. To find numerical values in the conditions of the iteration realization we use the goal function estimates obtained at the corresponding μ. Let us denote the goal function estimate with $q_t(\mu_k)$ which is calculated at the t-th iteration on the samples $z^{(2)}_{\bar{j},t}$ and $\tilde{z}^{(1)}_{\bar{j},t}$ of the local sample of size μ_k and as $q^{\pm}_t(\mu_k)$ – goal function estimate at the t-th iteration at the same μ_k calculated on samples $z^{(2)}_{\bar{j},t}$ and $\tilde{z}^{(1)}_{\pm}(\bar{j}_t, \hat{\alpha}_{1,t-1}, \ldots \hat{\alpha}_{i,t-1} \pm \Delta_{\alpha,i}, \ldots \hat{\alpha}_{m,t-1})$, i.e. when a certain increment $\Delta_{\alpha i} > 0$ is specified $\bar{j}_t \in \Omega_t \in \Omega$ for the parameter α_i estimate.

Let us fulfill the following condition of the iteration realization: the iteration of finding the current estimate $\hat{\alpha}_{i,t+1}$ is not performed and $\mu_{t\min}$ increases by 1 (a new pair of samples $z^{(2)}_{\bar{j},t}$ and $\tilde{z}^{(1)}_{\bar{j},t}$ is added to the local sample) in two cases:

– if at the current t-th iteration the estimate $q_t(\mu_{t\min})$ for the local sample size $\mu_{t\min}$ is «better» than both the values $q^+_t(\mu_{t\min})$ and $q^-_t(\mu_{t\min})$;

– if at the current t-th iteration the estimate $q_t(\mu_{t\min})$ for the local sample size $\mu_{t\min}$ is «worse» than the values $q^+_t(\mu_{t\min})$ and $q^-_t(\mu_{t\min})$ but at that $q^+_t(\mu_{t\min}) = q^-_t(\mu_{t\min})$.

After that the sample size increases by one ($\mu_{t\min}$ +1) and the above-mentioned conditions are verified again. If one of them is fulfilled, then μ increases by one again and so on right up to a certain value $\mu_{\max}$. If $\mu_{\max}$ is attained, the following iteration of the parameter α_i estimation is performed. If at the current μ the conditions are not met, then the next $(t+1)$-th iteration of the estimate $\hat{\alpha}_{i,t+1}$ forming for the parameter α_i is carried out.

In particular, when maximizing the goal function we can write the procedure of parameter α_i estimation in the following form

$$\hat{\alpha}_{i,t+1} = \hat{\alpha}_{i,t} - \lambda_{t+1}\beta_{i,t+1}(\mu_t);$$
$$\mu_t = \begin{cases} \mu_k + 1, \ if \ \left(q_t(\mu_k) < q_t^+(\mu_k) \wedge q_t^-(\mu_k)\right) \vee \left(q_t(\mu_k) > q_t^+(\mu_k) = q_t^-(\mu_k)\right), \\ \mu_{\max}, \ if \ \mu_k = \mu_{\max}, \\ \mu_k, \quad in \ the \ other \ case; \end{cases} \tag{19}$$
$$\beta_{i,t+1} = \begin{cases} 1, \ if \ \ q_t^+(\mu_t) < q_t^-(\mu_t) \wedge q_t(\mu_t) > \min\left(q_t^+(\mu_t), q_t^-(\mu_t)\right); \\ 0, \ if \ \ q_t^+(\mu_t) = q_t^-(\mu_t); \\ -1, \ if \ \ q_t^+(\mu_t) > q_t^-(\mu_t) \wedge q_t(\mu_t) > \min\left(q_t^+(\mu_t), q_t^-(\mu_t)\right). \end{cases}$$

Let us note that as t increases the value $\mu_{t\min}$ varies according to a certain prescribed rule that is defined by the problem to be solved, in particular, in the simplest case $\mu_{t\min} = const$.

In Fig. 5 experimental results obtained for the algorithm (19) realization are presented. In the experiment a real image of optical range with correlation radius equal to 5 steps on the sample grid was used. A parameter to be estimated was the parallel shift $\overline{h} = (10, 0.5)^T$. The shifted image was additionally noised by an independent centered Gaussian noise. The dependencies of μ_t as a function of the number of iterations, averaged on 50 realizations, are shown in Fig. 5,a. Here, the dependence 1 corresponds to signal variance-to-noise variance ratio g =100 and the dependence 2 – g =50. It is seen that for great errors of the estimate the sample size is small (for g =100 at t =10 it is equal to about 2 and at t =500 – to about 2.3) and it increases monotonously on average as the number of iterations increases (attaining about 6 at g =100 and t =2000). In Fig. 5,b plots of the estimation error ε_h versus the number of iterations are presented, where curve 1 corresponds to the results, obtained for the adaptive forming of μ_t, and curve 2 – at constant $\mu = \mu_m$, where $\mu_m = \sum_{k=1}^{2000} \mu_k$ – average sample size for t varying from 1 to 2000. The results are averaged on 250 realizations. It is obvious at small number of iterations there is a loss in estimation accuracy (at the 100th iteration it is about 5 per cent). It can be explained by high speed of the algorithm convergence with constant μ at the initial stage of estimation (due to a greater average of μ). However, at equal computational expenses (to the 2000th iteration) a gain in accuracy of about 2.4 times as large is guaranteed.

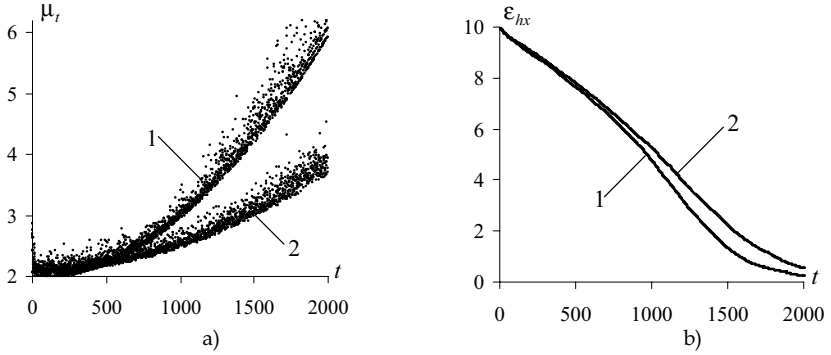


Fig. 5. An example of dependence of local sample size and estimate error versus the number of iterations

Thus, due to the fact that the proposed PGA with adaptive size of local sample facilitates the estimates vector recovery from local extremums of the goal function, it enables to significantly increase parameter estimate convergence speed in comparison with PGA with constant sample size at equal computational expenses. One more example of construction of PGA with adjustable local sample size is presented in work (Tashlinskii, 2003).

Let us note that if the problem of IIGD parameter estimation is a part of the problem of identification with a decision rule based on goal function values, then in order to achieve the required confidence probability of identification it may require large sample size μ_I that is not justified in the process of the estimates $\hat{\overline{\alpha}}$ convergence. In this case it is expedient to use adaptive adjustment of sample size and we can choose $\mu_{max} = \mu_I$ as its maximum. Then the attainment of μ_{max} will simultaneously mean the identification problem solution. An example of such a problem can be search of a fragment location on the reference image where a criterion of correspondence is in excess of a certain correlation sample coefficient value between the fragment and the reference image.

4. Structural optimization of pseudogradient algorithms

In practical problems of IIGDs estimation by means of pseudogradient procedures the required accuracy of parameter $\overline{\alpha}$ estimates is not obtained in the whole domain $\Omega_{\overline{\alpha}}$ of possible values $\overline{\alpha}$, but only in a certain subdomain bounded by an operating range of the procedures. This leads to the necessity of decomposing $\Omega_{\overline{\alpha}}$ into N subdomains $\Omega_{\overline{\alpha}}^{(i)}\left(\hat{\overline{\alpha}}_{0,i}\right)$, $i = \overline{1,N}$ corresponding to the operating range of the employed procedures where $\hat{\overline{\alpha}}_{0,i}$ – an initial approximation of the parameters for the procedure operating in the i-th subdomain. Let the procedure, operating in the subdomain which contains the sought vector $\overline{\alpha}_v$ of parameters, be called a V-procedure (from veritas - true) and the corresponding subdomain – V-subdomain. Subdomains that do not include $\overline{\alpha}_v$ are called P-subdomains (from pseudo – false) and the corresponding procedures – P-procedures. As the result of all these procedures operation N vectors $\hat{\overline{\alpha}}_i$ of IIGD parameter estimates are formed and the problem of determination of a V-subdomain among these estimates with required accuracy where the goal function attains its extremum arises.

4.1 The principle of pseudogradient procedure set control

In the problems of IIGDs estimation the number of subdomains can run up to dozens of thousands. Thus, bringing all the procedures operating in subdomains to the number of iterations that ensures the necessary accuracy of estimation requires great computational expenses. At such an approach the choice of the V-subdomain requires additional calculations. To reduce computational expenses the following principle of structural optimization can be used (pseudogradient procedures set control). At each step of the algorithm the priority of the current iteration realization is given to the procedure, having the least value of a certain penalty function ψ characterizing the level of the priority (Tashlinskii, 2006). Here, by «algorithm step» we imply a set of operations that includes performing of the current iteration by the procedure with the least penalty function, finding

a new value of the penalty function and obtaining a procedure with the least penalty function.

A characteristic property of such an approach is in the necessity to compare the «penalties» of the procedures which have performed different number of iterations. Studies have shown that when minimizing the goal function the following penalty functions satisfies such requirements

$$\psi_t^{(i)} = \sum_{k=1}^{t} \left(q_k^{(i)} - q_{\inf} \right), \; i = \overline{1,N},$$

where $q_k^{(i)}$ – goal function estimate at the k-th iteration; $q_{\inf} \le \inf\left\{ q_k^{(i)} \right\}$ – value which is less than the lower bound of the possible estimates set of the goal function. If the goal function is to be maximized then

$$\psi_t^{(i)} = \sum_{k=1}^{t} \left(q_{\sup} - q_k^{(i)} \right), \; i = \overline{1,N},$$

where $q_{\sup} \ge \sup\left\{ q_k^{(i)} \right\}$.

In the process of parameters estimates convergence to the optimal values the probabilistic properties of the goal function estimates are changing, which leads to the change of the probabilistic properties of the penalty function ψ. So when studying properties of ψ it is necessary to know its probability distribution density $w_t(\psi)$ at each iteration of estimation. At that, $w_t(\psi)$ depends on the local sample Z_t i.e. a rate of the correspondence (similarity) of the sets $\left\{ \tilde{z}_{j,t}^{(1)} \right\}$ and $\left\{ z_{j,t}^{(2)} \right\}$ involved in the local sample. It is expedient to use correlation sample coefficient ρ as a value characterizing this correspondence. For isotropic images ρ is a one-dimensional characteristic for any number of parameters to be estimated which simplifies calculations. Then for probability distributions of the penalty function increment $\Delta\psi$ at the t-th iteration we can write

$$w_t(\Delta\psi) = \int_{-1}^{1} w_t(\Delta\psi|\rho) w(\rho) \, d\rho, \tag{20}$$

where $w_t(\Delta\psi|\rho)$ – conditional density of increment; $w(\rho)$ – probability distribution density of the correlation coefficient. Let us note that for V-procedures $w_t(\Delta\psi|\rho)$ depends on the iteration number because ρ increases as the estimates vector $\hat{\bar{\alpha}}$ converges to the optimal values.

Without loss of generality, we can assume that ψ takes only positive values. Then, for calculation of the distribution density of ψ at the t-th iteration we can obtain the recurrent expression

$$w_t(\psi) = \int_0^{\infty} \int_{-1}^{1} w_{t-1}(\psi - \Delta\psi_t) w_t(\Delta\psi \,|\, \rho) w(\rho) \, d\Delta\psi \, d\rho. \tag{21}$$

For the P-procedures $w_t(\Delta\psi|\rho)$ does not depend on the iteration number. Then,

$$w_t(\psi) = \int_0^{\infty} w_{t-1}(\psi - \Delta\psi_t) w(\Delta\psi) \, d\Delta\psi, \tag{22}$$

where $w(\Delta\psi) = w_t(\Delta\psi|\rho = 0)$.

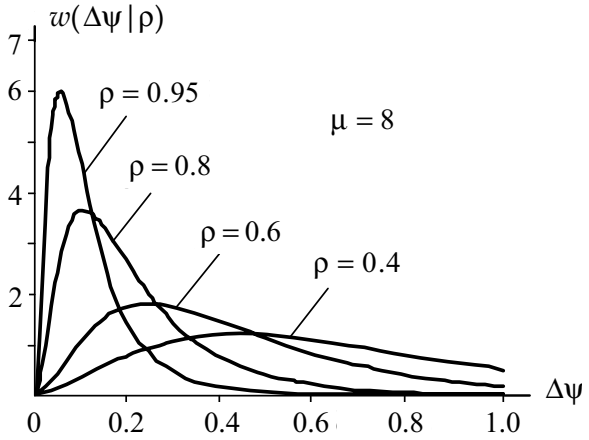
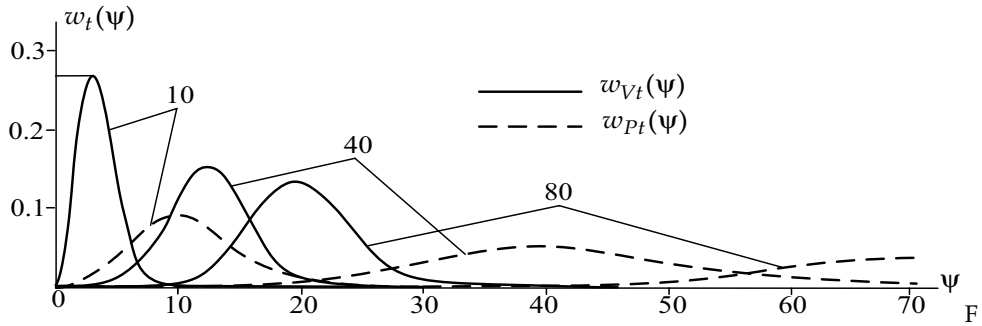


Fig. 6. Conditional distribution density of the penalty function increment

The expressions (20)–(22) enable to easily find the penalty function distribution density for the goal functions obtained in the second part. As an example, in Fig. 6 curves $w(\Delta\psi|\rho)$ of the increment $\Delta\psi$ of the interframe difference mean square at ρ =0.4, 0.6, 0.8, 0.95 are presented. Other parameters of calculation were the following: the images $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are Gaussian with correlation radius equal to 5 steps of the sample grid; the signal/noise ratio $g = 100$; the local sample size μ =4. In Fig. 7 the probability distribution densities of the interframe correlation sample coefficient for V-procedure ($w_{Vt}(\psi)$) and P-procedure ($w_{Pt}(\psi)$) at the number of iterations 10, 40 and 80 and μ = 7 are presented. From the plot it is seen that the area of intersection between $w_{Vt}(\psi)$ and $w_{Pt}(\psi)$ decreases sharply as the number of iterations increases, which facilitates reliable separation of the procedures of P- and V-type.



ig. 7. Penalty function distribution density at the interframe correlation sample coefficient