

Reduction of the computational complexity of stochastic gradient algorithms of image parameters estimation for a priori optimization of the local sample volume

A G Tashlinski¹, M G Tsaryov¹ and D G Kraus¹

¹Ulyanovsk State Technical University, Severny Venets str. 32, Ulyanovsk, Russia, 432027

Abstract. At stochastic gradient estimation of image parameters the estimates convergence character and computational expenses essentially depend on image samples local sample size used for obtaining the stochastic gradient. In the paper the possibility of a priori optimization of the volume of a local sample to minimize computational costs at geometrical images deformations estimation is considered. The minimum of the given computational costs for the conventional unit of expectation of the improvement of the evaluation is chosen as an optimization criterion. The block diagram of one of the algorithms and the examples of calculation results are presented.

1. Introduction

Recently, systems in which the initial information is a dynamic data array represented in the form of images have become widespread because such a presentation has visibility, compactness and information capacity.

The methods of image parameters estimation are based, as a rule, on four approaches. This is a comparison of image fragments, spatial-temporal filtering, morphological analysis [1], and analysis of the optical flow. For large-scale images, processing with stochastic gradient algorithms (SGA) [2-5], based on the analysis of the optical flow, is effective. In this case, the vector of estimates $\hat{\alpha}$ of the investigated parameters $\bar{\alpha}$ is formed iteratively:

$$\hat{\alpha}_t = \hat{\alpha}_{t-1} - \Lambda_t \bar{\beta}_t(Q), \quad (1)$$

where Λ_t – gain matrix, $\bar{\beta}_t$ – stochastic gradient of the objective function Q of the estimation quality, $t = \overline{1, T}$ – iteration number; $\hat{\alpha}_0$ – initial approximation of parameters. The algorithms are recurrent, have good accuracy characteristics and high speed of execution, do not require a preliminary evaluation of the images parameters, applicable to images with smoothly varying heterogeneity. The parameters estimated by the algorithm converge to the optimal values under rather weak conditions [6] and are resistant to impulse noise [7].

The study of the temporal dynamics of images requires analysis of the parameters of sequences of image frames, in particular, interframe geometric deformations of, for example, images $\mathbf{Z}^{(1)} = \{z_j^{(1)}\}$ and $\mathbf{Z}^{(2)} = \{z_j^{(2)}\}$, where $\bar{j} \in \Omega$, Ω – samples grid. As the objective function is chosen the mean square of the interframe difference in the estimation of interframe deformations under conditions when

it is possible to neglect the brightness distortions, or the sampling coefficient of interframe correlation at interframe brightness distortions close to linear [8].

The key problem is to increase the speed of the SGA. Various approaches are being explored to solve this problem. In particular, in [9] a procedure for stochastic gradient optimization of the second order in a linear time was proposed, in [10] to accelerate the optimization process using algorithms based on stochastic gradient descent the Nesterov moment method is applied, in [11] a convolutional neural network is used to estimate the geometric mismatch parameters between two images in accordance with a given geometric model (the affine model and the thin-plate spline transformation are studied), in [12] the acceleration of the stochastic gradient descent is achieved by taking into account the probability of smoothness of separate areas of the image. Also one of the approach is to reduce the volume μ of a two-dimensional local sample $Z_t = \{z_{jt}^{(2)}, \tilde{z}_{jt}^{(1)}\}$. It is used at each iteration of the estimation to find the stochastic gradient $\vec{\beta}(Q)$ of the objective function, where $z_{jt}^{(2)} \in \mathbf{Z}^{(2)}$, $\tilde{z}_{jt}^{(1)} = \tilde{\mathbf{Z}}^{(1)}$; $\tilde{\mathbf{Z}}^{(1)}$ – oversampled image $\mathbf{Z}^{(1)}$ using the current estimates $\hat{\alpha}_{t-1}$ of the interframe deformations parameters. But in so doing, the possibilities of a priori and a posteriori optimization of the local sample volume according to various optimality criteria have been poorly investigated. In this paper we consider the possibility of a priori optimization of the local sample volume by the criterion of minimum computational costs when estimating one parameter.

2. Optimizing the local sample volume

Suppose that, in accordance with the given error in the estimation of the parameter α , the mismatch $\varepsilon = \alpha_{ev} - \hat{\alpha}$ of the parameter estimate $\hat{\alpha}$ and its exact value α_{ev} should change from ε_{max} to ε_{min} . Consider the possibility of minimizing the computing costs of the SGA by optimizing the volume of the local sample for each iteration of the estimation for the given conditions. We use the following optimality criterion.

At each t -th iteration of the stochastic gradient estimation, we will search for such a volume μ_t of the local sample that provides the minimum of the computational cost per the conditional unit of the mathematical expectation of an parameter estimate $\Delta\alpha_t$ improvement

$$\mu_t = k \Big|_{\min_{\Delta\alpha_t} g(k)}, k = 1, 2, \dots, \quad (2)$$

where $g(k)$ – the computational cost of implementation by the algorithm of the t -th iteration for a local sample volume equal to k ; $g(k)/\Delta\alpha_t$ characterizes the computational costs, normalized to the conditional unit of the mathematical expectation $\Delta\alpha_t$ of an estimate improvement (an expression for the calculation $\Delta\alpha_t$ using relay type of stochastic gradient estimate sets out later in this paper).

Because at the estimation iterations the parameter mismatch consistently changes from $|\varepsilon_{max}|$ to $|\varepsilon_{min}|$, then for the T iterations the proposed criterion will provide the minimum total computational costs

$$G = \sum_{t=1}^T g(\mu_t), \quad (3)$$

where T – the number of iterations required to perform the condition $|\varepsilon_T| < |\varepsilon_{min}|$; ε_T – the mismatch of the parameter estimate $\hat{\alpha}$ and its exact value at the T -th iteration.

A detailed analysis of computational costs requires consideration not only the features and structure of the calculated ratio, but also many other influencing factors. These include the sampling time and conditions of images, the class of computing device, the time spent on the operations of addition, multiplication, division, access to memory, move data, and other auxiliary operations. Many of these factors depend on the specific image recording devices and the computers used. Therefore, in this

paper, the computational cost components will not be concretized. However, we will assume that the computational costs $g(\mu_t)$ of performing the SGA t -th iteration contain two components:

$$g(\mu_t) = g_{z_t} + g_o, \quad (4)$$

where g_{z_t} – computational cost for the formation of a local sample; g_o – other computing costs.

In this case, the cost of forming a local sample will be considered proportional to the volume μ of the local sample: $g_{z_t} = \mu \Delta g$, where Δg – computational costs for the formation of a local sample of a unit volume. Then

$$g(\mu_t) = \Delta g (c^{-1} + \mu_t), \quad (5)$$

where $c = \Delta g / g_o$ – the coefficient characterizing the proportion of the g_o when the volume of the local sample increases by one.

For relay SGA, the mathematical expectation of the improvement in the estimation $\Delta \alpha_t$ of the parameter under study by the t -th iteration can be found [13] by using the drift probability of the estimates [14]

$$\Delta \alpha_t = M[\varepsilon_{t+1} - \varepsilon_t] = (\varepsilon_t - \Lambda_t) \rho^- + \varepsilon_t \rho^o + (\varepsilon_t + \Lambda_t) \rho^+ = \varepsilon_t + \Lambda_t (\rho^+ - \rho^-), \quad (6)$$

where ρ^+ – the probability that, for a given mismatch ε , the estimate $\hat{\alpha}$ will change toward the exact value of the parameter (sign $\varepsilon_t = \text{sign } \Delta \alpha_{t+1}$); ρ^- – the probability that, for a given mismatch ε , the estimate $\hat{\alpha}$ will change away from the exact value of the parameter, that is sign $(\varepsilon_t) = -\text{sign } \Delta \alpha_{t+1}$; ρ^o – the probability that the estimate will not change ($\Delta \alpha_t = 0$). Obviously, these probabilities constitute exhaustive events: $\rho^+ + \rho^o + \rho^- = 1$. We also note that, in the strict sense, the drift probability ρ^+ characterizes not the probability of improving the estimate, but the probability of changing the estimate in the "right" direction.

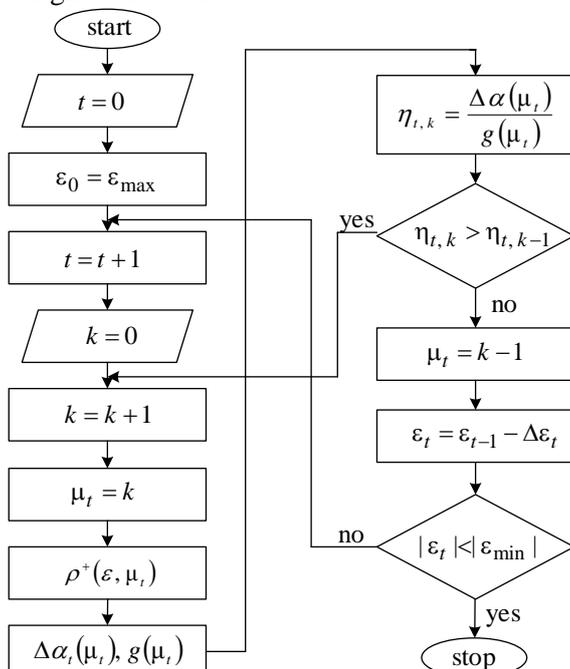


Figure 1. Block-diagram of the algorithm of the local sample volume optimization.

A block diagram of one of the possible algorithms for finding the optimal volume of a local sample are presented on figure 1. Here, for simplicity, it is assumed that $\rho^o = 0$, then $\rho^- = 1 - \rho^+$ and $\Delta \alpha_t = \Lambda_t (2\rho^+ - 1)$. To sequentially calculate the volume μ_t of the local sample at the t -th iteration,

$t = \overline{1, T}$, in the range of the deviation of the estimate from $|\varepsilon_{\max}|$ to $|\varepsilon_{\min}|$, the initial conditions are given $t=0$ and $\varepsilon_o = \varepsilon_{\max}$. Next, the volume μ_1 of the local sample is calculated, at which the minimum of the reduced computational costs (the minimum of the ratio $g(k)/\Delta\alpha_1$) is reached at the first iteration. Then, the local sample volume is estimated at the next iteration, i.e. is computed μ_2 and so on, until condition $|\varepsilon_t| < |\varepsilon_{\min}|$ is fulfilled.

3. Examples of calculating of the local sample volume

An example of computation results of optimal value of the local sample volume as a function of the mismatch is shown in figure 2. An interframe parallel image shift was evaluated. The value of the parameter c is chosen equal to 5%. Curve 1 corresponds to the noiseless images, and curve 2 corresponds to the signal-to-noise ratio $\sigma_x^2/\sigma_\theta^2 = 10$. It is assumed that the noise model of the researched images $\mathbf{Z}^{(1)}$ и $\mathbf{Z}^{(2)}$ is additive: $z_j = x_j + \theta_j$, where x_j – image with dispersion σ_x^2 , θ_j – independent Gaussian noise with zero mathematical expectation and variance σ_θ^2 .

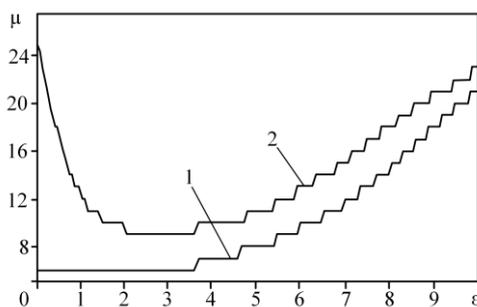


Figure 2. Relation the mismatch with the optimal of the local sample volume.

For the conditions corresponding to curves 1 and 2, the table 1 shows the results of the experiment. They show the loss in computing costs when using a constant volume of local sample ($\mu = const$) in comparison with the case of using the optimal volume of local sample. When specifying in SGA $\mu = const$, the value of μ corresponded to the average value μ_{avg} of the optimal sample size, and $\mu_{\text{avg}} - 2$, $\mu_{\text{avg}} - 1$, $\mu_{\text{avg}} + 1$ и $\mu_{\text{avg}} + 2$.

	$\mu_{\text{avg}} - 2$	$\mu_{\text{avg}} - 1$	μ_{avg}	$\mu_{\text{avg}} + 1$	$\mu_{\text{avg}} + 2$
Curve 1	4.1	3.9	3.9	4.5	4.9
Curve 2	2.1	1.7	1.8	2.4	2.9

4. Conclusion

The approach to increasing the rate of algorithms for stochastic gradient estimation of image parameters is considered on the example of inter-frame deformations estimation. The purpose is achieved by optimizing the size of the two-dimensional local sample used at each iteration of the estimation to determine the stochastic gradient. A priori optimization based on the criterion of minimum computational costs for the case of estimating one parameter is used. At the same time at each iteration of the estimation, the local sample size providing a minimum of computational costs for the conventional unit of the mathematical expectation of the parameter estimate improvement is determined. It is shown that, as the number of iterations increases, the mismatch modulus of the estimate and the exact value of the parameter decreases, the proposed approach provides a minimum of total computational costs. To determine the mathematical expectation of an improvement of the

studied parameter estimate, the probabilities of drift estimates (the probability of changing the estimates towards the exact value of the parameter and from it) are used.

The carried out modeling for one of the possible algorithms realizing the proposed approach confirmed the set purpose. So, for the given example of results, the gain in computational costs in comparison with the situation of using the constant volume of the local sample amounted to no less than 3.9% in the absence of noise, and not less than 1.8% for a signal-to-noise ratio equal to 10 (in variance). Thus, the proposed approach for algorithms of stochastic gradient estimation of image parameters makes it possible to determine the optimal size of the local sample for each estimation iteration, which ensures the minimization of computational costs.

5. References

- [1] Rubi A Yu, Lebedev M A, Vizilter Yu V and Vygolov O V 2016 Morphological image filtering based on guided contrasting *Computer Optics* **40(1)** 73-79
- [2] Su H R and Lai S H 2015 Non-rigid registration of images with geometric and photometric deformation by using local affine Fourier-moment matching *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 2874-2882
- [3] Moritz P, Nishihara R and Jordan M I 2016 A linearly-convergent stochastic L-BFGS algorithm *Proc. of the 19th Int. Conf. on Artificial Intelligence and Statistics, AISTATS* pp 249-258
- [4] Borisova I V, Legkiy V N and Kravets S A 2017 Application of the gradient orientation for systems of automatic target detection *Computer Optics* **41(6)** 931-937
- [5] Taslinskii A G 2008 Optimization of goal function pseudogradient in the problem of interframe geometrical deformations estimation *Pattern Recognition Techniques, Technology and Applications* (Vienna: I-Tech) pp 249-280
- [6] Tsyppkin Ya Z 1995 *Information theory of identification* (Moscow: Fizmatlit) p 336 (in Russian)
- [7] Tashlinskii A G 2007 Pseudogradient estimation of digital images interframe geometrical deformations *Vision Systems: Segmentation & Pattern Recognition* (Vienna: I Tech Education and Publishing) pp 465-494
- [8] Tashlinskii A G 2008 The specifics of pseudogradient estimation of geometric deformations in image sequences *Pattern Recognition and Image Analysis* **18(4)** 701-706
- [9] Agarwal N, Bullins B and Hazan E 2017 Second order stochastic optimization for machine learning in linear time *Journal of Machine Learning Research* **18(116)** 1-40
- [10] Allen-Zhu Zeyuan 2017 Katyusha: the first direct acceleration of stochastic gradient methods *Proc. of the 49th Annual ACM SIGACT Symposium on Theory of Computing* pp 1200-1205
- [11] Rocco I, Arandjelovic R and Sivic J 2016 Convolutional neural network architecture for geometric matching *Proc. CVPR* **2** 6148-6157
- [12] Allen-Zhu Zeyuan, Richt'arik Peter, Qu Zheng and Yuan Yang 2016 Even faster accelerated coordinate descent using non-uniform sampling *Proc. of the 33rd Int. Conf. on ICML* **48(4)** 1110-1119
- [13] Tashlinskii A G and Tikhonov V O 2001 The method for analyzing the error of pseudo-gradient measurement of multidimensional processes parameters *Izvestiya vuzov: Radioelektronika* **44(9)** 75-80 (in Russian)
- [14] Tashlinskii A G and Voronov I V 2014 The probability of demolition of estimates of parameters of interframe geometric deformations of images under pseudo-gradient measurement *Izvestiya of the Samara Scientific Center of the RAS* **16 N6(2)** 612-615 (in Russian)

Acknowledgments

The reported study was funded by RFBR and Government of Ulyanovsk Region according to the research projects 16-01-00276 and 18-41-730006.