The Analysis of the Morphology of Nanostructures in Noise

Lev V. Novikov
Institute for analytical instrumentation of RAS
St. Petersburg, Russia
novik38@mail.ru

Victoria V. Kurkina
St. Petersburg state Institute of technology (technical University)
St. Petersburg, Russia
victoria.kurkina@gmail.com

***Abstract*— A new noise-tolerant algorithm for estimating the position and intensity of nanoscale image peaks is proposed. The method is based on sequential processing of images by rows and columns to detect the beginning, maximum and end of the peaks by derivatives at three points of the sliding data window. The coordinates of the points found in this way form clusters of peaks, which are further analyzed by the cluster analysis method.**

Unlike traditional approaches to surface structure analysis, the proposed method does not require noise pre-filtering. With a signal/noise ratio of up to three, it provides a probability of detecting at least equal to one with zero probability of missing and false peak detection.

***Keywords*— morphology of nanostructures; image processing; peak detection; nanoscale images**

One of the problems solved in the theory of image processing is to obtain data on the surface morphology, for example, the morphology of nanostructures used in various applications from fingerprint recognition and diagnostics of printed circuit boards [1] to assess the surface roughness, determining the size of nanoparticles and the quality of biosimulator in the pharmaceutical industry [2].

The reliability of decisions on the evaluation of morphological parameters, in particular nanostructures, depends on the accuracy of the estimates of these parameters in conditions of strong noise and other disturbances. This work is devoted to the assessment of the intensity and position of the peaks of images typical for complex studies using atomic force, tunnel or raster electron microscopy.

The traditional approach in the analysis of surface structure in the presence of noise is based on preliminary noise filtration using some well-known algorithms. However, this operation inevitably leads to the distortion of the surface characters – the violation of the shape of the peaks, reducing their intensity and, sometimes, the position of the maxima. The proposed algorithm does not require pre-filtering. It is based on the method of searching for the beginning, maximum and end of the peaks by using the derivative at three points of the sliding data window proposed in the work [3].

The searching of images of nano-objects maxima is performed in two stages. On the first one-by detecting extrema on rows and columns of the data matrix the clusters of points in the peak zone are formed. Then, using hierarchical cluster analysis, these clusters are localized and their centers are determined, the size and location of which are taken as the peak parameters in the image.

An image containing a set of peaks can be represented as a matrix **Z**, the elements of which $z(i,j)$ are determined by the formula

$$z(i,j)=\sum_{l=1}^{L} A_l f_l\left((i-i_l),(j-j_l)\right)+\rho(i,j)$$

where A_l – intensity; f_l – function of form; i_l, j_l – position of l-th peak; L – the number of peaks; $\rho(i,j)$ – additive noise; $i=\overline{(1:I)}$, $j=\overline{(1:J)}$.

The peak model, which is a Gaussian peak with an ellipsoidal cross-section by a horizontal plane is used in this paper to test the algorithm:

$$f_l(.)=A_l\,exp\left\{-\frac{r_l^2}{\mu_l^2}\right\}, \tag{1}$$

where $r_l^2=\left((i-i_l)^2+(j-j_l)^2\right)\Big/A_l^2-\frac{\mu_l^2}{\beta_l^2}\left((i-i_l)\cos\,\alpha_l-(j-j_l)\sin\,\alpha_l\right)^2\Big/A_l^2$

Here, $2\mu_l$ is the focal distance and $2\beta_l$ – the semimajor axis of the ellipse $\beta_l>\mu_l$, α_l – the angle of the axis of the ellipse in the plane (i,j), $A_l=\sqrt{\beta_l^2-\mu_l^2}$

Each j-th column of the matrix **Z** forms a I-dimensional column-vector

$$Z_j=\left[z_j(i)\right]$$

with elements $z_j(i)$.

Each i-th row of the matrix **Z** forms a J-dimensional row-vector

$$Z_i=\left[z_i(j)\right]$$

with elements $z_i(j)$.

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

The algorithm for detecting the position (i_l, j_l) and the amplitude A_l of the peaks of the image is carried out by sequential processing of the columns and rows of the matrix **Z**, followed by analysis of the data to clarify the required parameters.

When processing on columns Z_j the search of inflections (ridges) in each column j of a matrix **Z** is made. For this purpose, the approach described in detail in [3] for the one-dimensional case is used. The essence of the proposed algorithm is as follows:

Let's form the current sample **y** from $K+1$ samples with the center at the point i of the vector Z_j

$$\mathbf{y} = \left[y(k)\right] = \left(z_j\left(i-K/2\right),...,z_j\left(i\right),...,z_j\left(i+K/2\right)\right) ,$$

where $k = \overline{(1:K)}$ and K is chosen always an even number from the condition $0.5 \div 0.7$ of the peak width at half the height.

The points in the window **y** is approximated by a second degree polynomial (parabola) $\tilde{y}$ on the grid $k = \left(-K/2,...,0,...,K/2\right)$:

$$\tilde{y}(k) = p_1 k^2 + p_2 k + p_3 ,$$

where p_1, p_2, p_3 the coefficients of the polynomial.

Let's choose three points t_1, t_0, t_2 in the interval $\left[-\frac{K}{2},\frac{K}{2}\right]$, such that $t_1 = -\left\lceil \frac{K}{4} \right\rceil, t_0 = 0, t_2 = \left\lceil \frac{K}{4} \right\rceil$, where $\left\lceil \frac{K}{4} \right\rceil$ – is an integer greater than or equal $\frac{K}{4}$.

The value of the first derivative of the parabola at the point t_1 will be equal to

$$d_1 = 2p_1 t_1 + p_2 \tag{2}$$

and in point t_2:

$$d_2 = 2p_1 t_2 + p_2 \ . \tag{3}$$

The value of the second derivative of the parabola is $2p_1$.

The value of the parabola at the point $k = 0$ is equal to the weighted-average value in the sliding sample of the initial data at the point i:

$$\bar{Z}_j\left(i\right) = \tilde{y}\left(0\right) = p_3 . \tag{4}$$

Using the formulas (2), (3), (4) the algorithms of image extrema detection (the beginning, maxima and minima of ridges and peaks) are formed.

Let us denote the numbers of the lines in which the beginnings of the ridges are found, as I_{minmax}.

If the beginning of the ridges is detected, their maxima are determined. Let us denote the numbers of the rows in which the ridge maxima are found, as I_{max}.

Further, under the condition of detection of maxima, the minima can be found and the criterion of overlapping of neighboring ridges can be determined.

Let us denote the numbers of the lines in which the ends of the ridges are found, as I_{maxmin}.

We form three new matrices Z_{minmax}, Z_{max}, and Z_{maxmin} consisting of all the columns of the matrix Z and rows with numbers I_{minmax}, i.e. $Z_{minmax} = Z\left(I_{minmax}, 1:J\right)$; I_{max}, i.e. $Z_{max} = Z\left(I_{max}, 1:J\right)$ and I_{maxmin}, i.e. $Z_{maxmin} = Z\left(I_{maxmin}, 1:J\right)$.

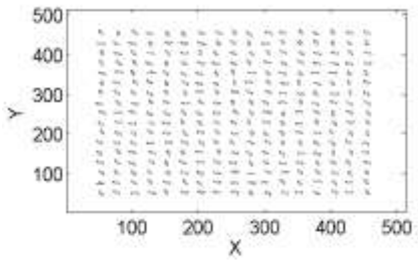
Perform the analysis of the rows I_{minmax}, I_{max} and I_{maxmin} of the matrices Z_{minmax}, Z_{max} and Z_{maxmin} respectively with the use of algorithms (4)-(8) in order to define the column numbers J_{minmax}, J_{max}, and J_{maxmin} in which the beginning, maxima and ends of ranges are detected.

As a result of the above analysis, we obtain in the form of a table the coordinates of the peaks beginning $\left[I_{minmax}, J_{minmax}\right]$, their maxima – $\left[I_{max}, J_{max}\right]$ and ends $\left[I_{maxmin}, J_{maxmin}\right]$.

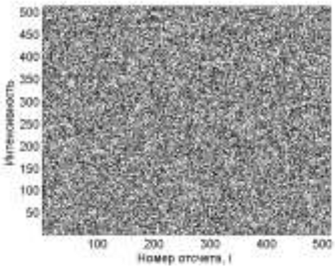
The coordinates of the beginning and end of the peaks form the table of minima of the analyzed surface $\left[I_{min}, J_{min}\right]$.

The amplitudes $\hat{A}\left(I_{max}, J_{max}\right)$ of the detected peaks are estimated by the maximum point of the parabola at the top of the ridge. If the shape of the peaks is asymmetric, several peaks forming a cluster are found in the neighborhood of their vertices even in the absence of noise. Each cluster is localized using known cluster analysis algorithms. The criterion of belonging to the cluster is the distance between the points, which should be less than the minimum distance between the peaks of the analyzed surface, set by a priori data. The average weighted value for each coordinate is taken as an estimate of the cluster position, and the maximum intensity of the peaks included in the cluster is taken as the its intensity.

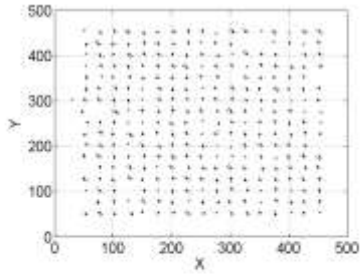
Verification of the effectiveness of the algorithm was carried out according to the model spectrum (formula (1)) with the addition of white noise. The position and intensity of the peaks at different noise levels were estimated. The Fig. 1 shows: a) the original image, b) a noisy image and (c) clusters of peaks of the image and their vertices obtained on the basis of the proposed algorithm. The results were compared with the image extremum detection program – extrema2.m used in the an expansion in the empirical modes (BEMD) [4, 5], with pre-filtering noise on the basis of multiwavelets (see Table).



a)



б)



в)

Fig. 1. The image model a) without noise, b) in the presence of noise with a signal-to-noise ratio equal to 3, and c) - the same model after processing by the proposed algorithm with the location of the peak peaks (•).

TABLE I. THE PROBABILITY OF A CORRECT AND FALSE DETECTION FOR DIFFERENT RATIOS OF THE SIGNAL TO NOISE (S / N): FOR 1- THE METHOD OF SEARCHING FOR EXTREMA, GIVEN IN [4,5], AND 2 - FOR THE PROPOSED METHOD

Ratio (s/n) / **Parameter**	3		10		20		100	
	1	2	1	2	1	2	1	2
The probability of correct detection	0.907	0.931	0.976	0.999	0.970	1.000	0.986	1.000
The probability of false detection	0.727	0.034	0.028	0.002	0.020	0.000	0.013	0.000

The probability of correct detection was calculated as the ratio of the number of detected peaks for all ten realizations, the position of which differs from the model by no more than four points, to the total number of model peaks. The probability of false detection was calculated as the ratio of the number of detected peaks for all ten realizations not specified in the model to the total number of model peaks. Comparison of columns 1) and 2) of the table shows a clear advantage of the proposed algorithm, especially in terms of the probability of false detection.

The proposed detection algorithm makes it possible to increase the reliability of peak detection, thanks to a low levels of estimates errors and the probability of false detection at small signal-to-noise ratios. This makes it possible to significantly lower the sensitivity threshold and expand the range of the parameters being evaluated.

REFERENCES

[1] Kaur B.,Sangeet Kaur P. Applications of Mathematical Morphology in Image Processing: A Review. International Journal of Electronics & Communication Technology. 2013. V.4, no. 3. Pp.15-17.

[2] Reichelt W.N., Kaineder A., Brillmann M. et al. High throughput inclusion body sizing: Nano particle tracking analysis. Biotechnol. J. 2017. No. 12. Pp.1-11. DOI: 10.1002/biot.201600471

[3] Novikov L.V.H, Kurkina V.V. Peak detection by differentiation at three points of the sliding data window. Trudy rudy mezhdunarodnoy nauchnoy konferentsii “Matematicheskiye metody v tekhnike i tekhnologiyakh” [Proc. of the Int. Sci. Conf. “Mathematical methods in engineering and technology”]. Saint Petersburg. 2017, vol.6, pp. 68-73 (in Russian).

[4] Weihua Dong, Xian’en Li, Xiangguo Lin, Zhilin Li. A Bidimensional Empirical Mode Decomposition Method for Fusion of Multispectral and Panchromatic Remote Sensing Images. Remote Sens. 2014. V. 6. Pp. 8446-8467.

[5] https://www.mathworks.com/matlabcentral/fileexchange/?utf8=%E2%9C%93&term=extrema2.m (assessed 06.06.2018)