Multi-object tracking for motion pattern analysis in biomedical research

Roman Matrosov [1], Mikhail Bogachev [1], Asya Lyanova [1], Aleksandr Sinitca [1], Nikita Pyko [1], Airat Kayumov [2]

[1] Radio Systems Dept., St. Petersburg Electrotechnical University, 197022, St. Petersburg, Russia – rmmatrosov@stud.etu.ru, mibogachev@etu.ru, ailianova@etu.ru, amsinitca@etu.ru, nspyko@etu.ru
[2] Institute of Fundamental Medicine and Biology, Kazan Federal University, 420008, Kazan, Russia – kairatr@yandex.ru

Keywords: computer vision, local edge density, multi-threshold analysis, trajectory extraction, Kalman filtering

Abstract

Motion structure analysis is a ubiquitous tool for the evaluation of the dynamical properties of various objects. Automation of image analysis allows not only speeding up the research, but also eliminating the subjective bias characteristic inevitable under manual data processing conditions. The complexity of this task increases dramatically when not one, but several objects of interest are simultaneously tracked using the same videos. In turn, this requires specialized algorithms that uniquely associate selected objects with tracks. Here we present an original algorithm for multiple objects tracking based on computer vision algorithms developed for the image segmentation, as well as subsequent adaptive object selection and tracking. This approach is particularly tailored to the complex motion patterns typically exhibited by natural movements of the living objects of interest in biomedical and ecological research from microscales (e.g., microbes or cells) to macroscales (e.g. animals in laboratory or natural living conditions).The data processing pipeline implementing the proposed algorithms, as well as their validation using simulated data are available at https://github.com/Digiratory/StatTools/blob/main/research/kalman_filter.ipynb.

1. Introduction

Motion pattern analysis is a ubiquitous tool for the evaluation of dynamical properties of various objects of interest in a variety of biomedical, pharmacological, environmental and other problems. Common objects of interest include tracking of individual microbes and cells, as well as macro-organisms such as animals, fishes, etc. Motion pattern are studied both under normal living conditions (e.g. in ecological monitoring), as well as under specific laboratory test conditions where experimental animals are subjected to treatment with perspective drugs; bio-indicatory organisms such as crayfish, fish and shellfish, which are widely used to assess water quality in both engineered reservoirs and natural habitats; microscopic organisms such as cells, bacteria and viruses, which are subjected to the effects of various substances on their vital activity, such as perspective antimicrobials, biocides, chemotherapeutic agents etc.

Recent advancements in video recording and image processing technologies bring this research to a new level, providing opportunities to record long-term observations, analyze them in a rather automated manner by extracting movement patterns and analyzing them to evaluate their normal behavior as well as stress response patterns. Automation of image analysis not only speeds up the research, but also eliminates subjective bias common for manual data processing.

In this regard, automated video information processing systems are becoming more widespread, and require dedicated instruments at all stages from image segmentation to the selection of objects of interest and their movement patterns analysis. The latter is largely based on the analysis of their trajectories, which imply using some tracking technologies to extract motion patterns from video recordings. The complexity of this problem increases drastically when not a single, but rather multiple objects of interest are being tracked simultaneously from the same video recordings, which in turn required specific algorithms that associate selected objects with tracks uniquely.

In this work, we present an original algorithm for multi-object tracking based on the in-house developed computer vision algorithms for the image segmentation, as well as subsequent adaptive object selection and tracking.

2. Image segmentation and object selection

In computer vision, one of the universal ways to identify potential objects of interest in images is to identify sharp transitions or edges, corresponding to object boundaries or its structural elements. When there are multiple objects of interest, including large conglomerates where they partially overlap in the images, such as microbial colonies, groups of animals etc., these subpopulations are typically characterized by high edge densities, in marked contrast to empty surface areas where no characteristic objects can be observed, allowing to use the local edge density for their identification based solely on its structural features (Sinitca, 2023a). Recently, we have successfully applied this concept to the segmentation of microbial and cell populations (Sinitca, 2023b), as well as ecological zones in remote sensing images (Sinitca, 2023c). In this approach, the smoothed local edge density is used as a virtual channel which is compared against certain threshold(s) for either single- or multi-class segmentation, respectively. For the threshold selection, we employ an *a posteriori* approach with multiple thresholds being tested for the entire image, and those corresponding to the expected image shapes are chosen for each individual objects (Volkov, 2020). While selection criteria generally depend on the shapes of the objects of interest, compact objects with typically circular shapes can be easily selected based on the minimization of the perimeter/area ratio (Bogachev, 2019) or inverse circularity (Mendoza and Lu, 2015), since circle exhibits minimum perimeter for a given area, respectively.

3. Association of selected objects with trajectories

While the above procedures lead to compact objects selection in individual frames, the next problem is their correct association between different frames to provide motion trajectories of single objects. Correct track association is largely complicated due to the partial overlap between objects which prevents their resolution in individual frames, crossing of trajectories, randomness of the motion patterns etc.

In a more general scenario, missing and duplicated points, as well as anomalous measurement errors, require adequate algorithms to extrapolate, interpolate and filter trajectories.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-189-2025 | © Author(s) 2025. CC BY 4.0 License.

Tailored design of the above algorithms is associated with the knowledge of the relevant imitation model capable of simulating tracks with similar statistical properties as the observational ones from purely independent (while noise) data.

In recent decades, it has been shown that various biological objects in both natural and laboratory environments exhibit long-term correlated motion patterns, leading to power laws in their fluctuation functions (*FF*) determined as the standard deviations of the object's random displacements in a finite time window *s*

$$FF(s) = \sigma\left[\frac{1}{k}\sum_k (X_{k+s} - X_k)\right] \sim s^H,$$

where H is known as the Hurst exponent. In a more general scenario such as motion patterns in restricted environments, H may not be a single value, but rather update with *s*, thus leading to the variable slope $H(s)$ (Bogachev, 2023).

The latter, according to the Wiener-Khinchine theorem, also leads to power laws in the respective power spectral densities

$$S(f) \sim f^{-\beta}, \beta = 2H - 1,$$

closely reminiscent of those observed in colored noise series.

Association of selected points with tracks is commonly based on the minimal distance from the expected track position that in turn can be obtained by the extrapolation of the previously observed track positions. The optimal extrapolation rule for a long-term power-law correlated random series is governed by (Kasdin, 1995)

$$\mathrm{x_n} = \sum_{k=1}^{n} -a_k x_{n-k},$$

where $a_k = \left(k - 1 - \frac{\beta}{2}\right) * \frac{a_{k-1}}{k}$, n is the filter length, which is theoretically unbounded, while in most practical scenarios chosen simply large enough to achieve the desired accuracy.

With the knowledge of the coefficients a_k, one can easily obtain the coefficients of the tailored Kalman filter. For example, the 2nd order Kalman filter is described by

$$\begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \cdot \begin{bmatrix} x_{k-1} \\ \dot{x}_{k-1} \end{bmatrix}$$

where $\dot{x}_k = x_k - x_{k-1}$.
The above can be rewritten in the scalar form

$$\begin{cases} x_k = F_{11}x_{k-1} + F_{12}(x_{k-1} - x_{k-2}) \\ x_k - x_{k-1} = F_{21}x_{k-1} + F_{22}(x_{k-1} - x_{k-2}) \end{cases}$$

$$\begin{cases} x_k = (F_{11} + F_{12})x_{k-1} - F_{12}x_{k-2} \\ x_k = (1 + F_{21} + F_{22})x_{k-1} - F_{22}x_{k-2} \end{cases}$$

In the first approximation, the autoregressive filter can be also described by the first two coefficients

$$x_k = -a_1 x_{k-1} - a_2 x_{k-2}.$$

Next, by substituting

$$\begin{cases} F_{11} + F_{12} = -a_1 \\ -F_{12} = -a_2 \end{cases} \Rightarrow \begin{cases} F_{11} = -a_1 - a_2 \\ F_{12} = a_2 \end{cases}.$$

And

$$\begin{cases} 1 + F_{21} + F_{22} = -a_1 \\ -F_{22} = -a_2 \end{cases} \Rightarrow \begin{cases} F_{21} = -1 - a_1 - a_2 \\ F_{22} = a_2 \end{cases}$$

altogether finally yields

$$F = \begin{bmatrix} -a_1 - a_2 & a_2 \\ -1 - a_1 - a_2 & a_2 \end{bmatrix},$$

where F is the transition matrix of the tailored Kalman filter.

Calculation of the 3rd order Kalman filter

$$\mathrm{F} = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix}$$

$$\begin{bmatrix} x_k \\ \dot{x}_k \\ \ddot{x}_k \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \cdot \begin{bmatrix} x_k \\ \dot{x}_{k-1} \\ \ddot{x}_{k-1} \end{bmatrix},$$

where $\dot{x}_k = x_k - x_{k-1}, \ddot{x}_k = \dot{x}_k - \dot{x}_{k-1}$.

It's obvious that $\ddot{x}_k = x_k - 2x_{k-1} + x_{k-2}$.

Let's perform matrix multiplication:

$$\begin{cases} x_k = F_{11}x_{k-1} + F_{12}(x_{k-1} - x_{k-2}) + \\ +F_{13}(x_{k-1} - 2x_{k-2} + x_{k-3}) \\ x_k - x_{k-1} = F_{21}x_{k-1} + F_{22}(x_{k-1} - x_{k-2}) + \\ +F_{23}(x_{k-1} - 2x_{k-2} + x_{k-3}) \\ x_k - 2x_{k-1} + x_{k-2} = F_{31}x_{k-1} + \\ +F_{32}(x_{k-1} - x_{k-2}) + F_{33}(x_{k-1} - 2x_{k-2} + x_{k-3}) \end{cases}$$

$$\begin{cases} x_k = (F_{11} + F_{12} + F_{13})x_{k-1} + \\ +(-F_{12} - 2F_{13})x_{k-2} + F_{13}x_{k-3} \\ x_k = (1 + F_{21} + F_{22} + F_{23})x_{k-1} + \\ +(-F_{22} - 2F_{23})x_{k-2} + F_{23}x_{k-3} \\ x_k = (2 + F_{31} + F_{32} + F_{33})x_{k-1} + \\ +(-1 - F_{32} - 2F_{33})x_{k-2} + F_{33}x_{k-3} \end{cases}$$

The process model is generated by a third-order AR filter.:

$$x_k = -a_1 x_{k-1} - a_2 x_{k-2} - a_3 x_{k-3}.$$

Let's compare the expression above sequentially with the three lines of equations.

$$F_{13} = -a_3;\ F_{23} = -a_3;\ F_{33} = -a_3.$$

Comparison with the first equation yields the following:

$$\begin{cases} F_{11} + F_{12} + F_{13} = -a_1 \\ -F_{12} - 2F_{13} = -a_2 \end{cases} \Rightarrow \begin{cases} F_{11} + F_{12} - a_3 = -a_1 \\ -F_{12} + 2a_3 = -a_2 \end{cases} \Rightarrow$$

$$\begin{cases} F_{11} = -a_1 + a_3 - F_{12} \\ -F_{12} = -2a_3 - a_2 \end{cases} \Rightarrow \begin{cases} F_{11} = -a_1 - a_2 - a_3 \\ F_{12} = a_2 + 2a_3 \end{cases}.$$

Comparison with the second equation yields the following:

$$\begin{cases} 1 + F_{21} + F_{22} + F_{23} = -a_1 \\ -F_{22} - 2F_{23} = -a_2 \end{cases} \Rightarrow \begin{cases} F_{21} = -1 - a_1 - a_2 - a_3 \\ F_{22} = a_2 + 2a_3 \end{cases}.$$

Comparison with the third equation yields the following:

$$\begin{cases} 2 + F_{31} + F_{32} + F_{33} = -a_1 \\ -1 - F_{32} - 2F_{33} = -a_2 \end{cases} \Longrightarrow$$

$$\begin{cases} F_{31} = -2 - F_{32} - a_1 + a_3 \\ -1 - F_{32} = -a_2 - 2a_3 \end{cases} \Longrightarrow$$

$$\begin{cases} F_{31} = -1 - a_1 - a_2 - a_3 \\ F_{32} = -1 + a_2 + 2a_3 \end{cases}.$$

altogether finally yields

$$F = \begin{bmatrix} -a_1 - a_2 - a_3 & a_2 + 2a_3 & -a_3 \\ -1 - a_1 - a_2 - a_3 & a_2 + 2a_3 & -a_3 \\ -1 - a_1 - a_2 - a_3 & -1 + a_2 + 2a_3 & -a_3 \end{bmatrix}.$$

4. Simulation study

Figures 1-3 show the results of simulation data series with specified values of the Hurst index and the results of their processing using Kalman filters of the 1st, 2nd and 3rd orders.

We simulated Gaussian data series with zero mean and unit variance exhibiting linear long-term correlations (LTC) with $FF(s) \sim s^H$ typically occurring in complex systems, where H is the Hurst exponent. We used the algorithm by (Kasdin, 1995) with $0.5 \le H \le 1.5$ with step $H = 0.1$ to generate $N = 10$ random series of size $L = 2^{12}$. Two types of data disturbances were simulated, the first one represented by additive white Gaussian noise (AWGN) with signal-to-noise ratios (SNR) defined as the ratio between the standard deviations of the signal and the noise, respectively, in the range [-20…10] dB. The second type of disturbance considers missing data (gaps) occurring at random points in the series (Poisson flow) with exponential gap length with two free parameters, the average gap length and the average gap rate, respectively.

Figure 1 illustrates the reconstruction of original data series from the disturbed measurements described above (with noise of the left and gaps on the right, top to bottom: simulated, disturbed and reconstructed data, respectively). The figure shows that, as expected, increasing the order of the Kalman filter reduces the noise level in the reconstructed signal.

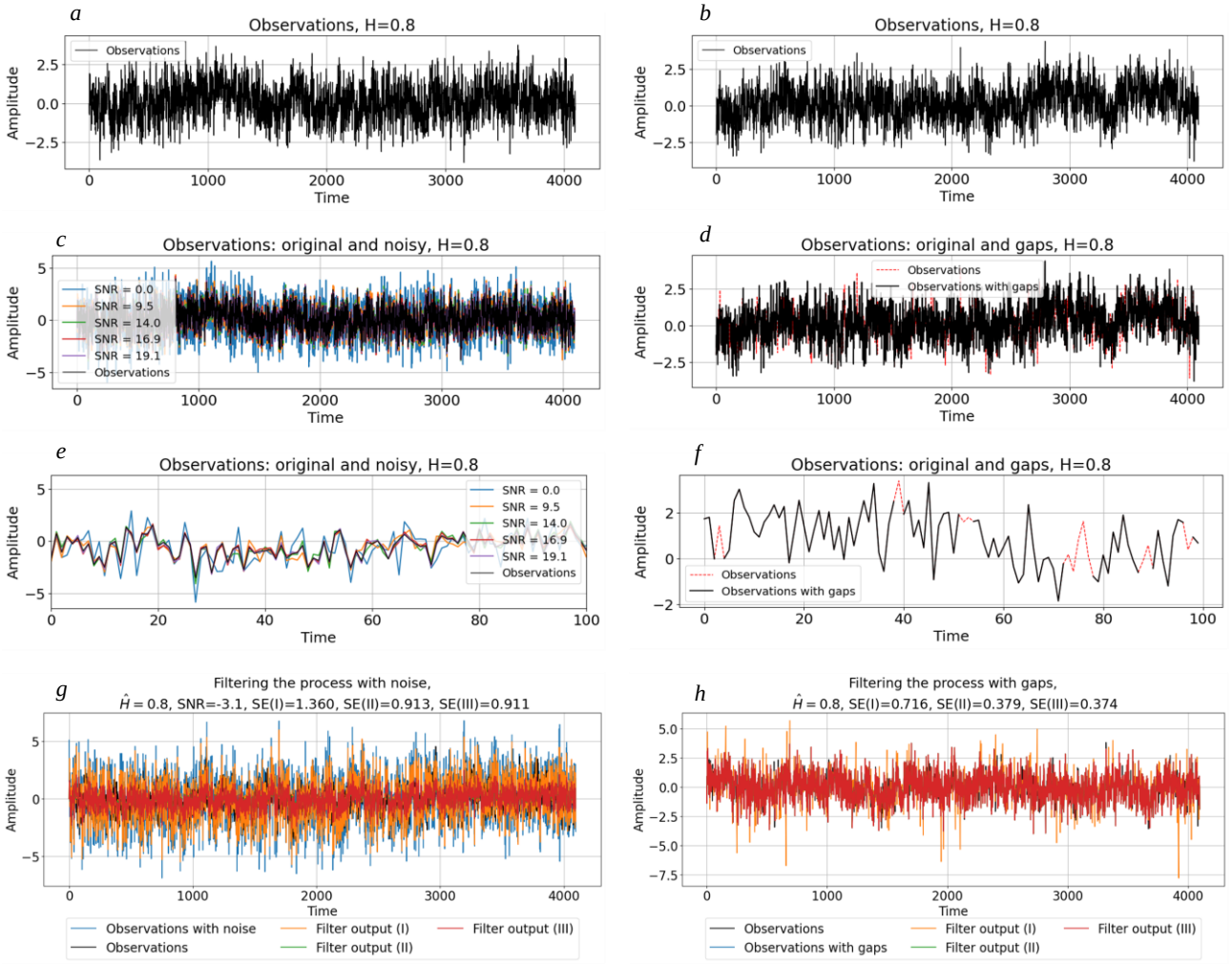

Figure 1. Data reconstruction using Kalman filters (I, II and III order) in the presence of gaps of additive noise of a given intensity (left column) and in the presence of gaps (right column). The left panel shows, from top to bottom, an example of the implementation of a series of data (*a*), a series of data with introduced interference (additive Gaussian noise of varying intensity) (*c*), a fragment of a series of data with omissions on an enlarged scale (*e*), the results of restoration (*g*), respectively, by filters of 1-3 orders. The right panel shows, from top to bottom, an example of the implementation of a series of data (*b*), a series of data with introduced interference (gaps) (*d*), a fragment of a series of data with noise on an enlarged scale (*f*), and filtering results using filters of different orders (*h*).

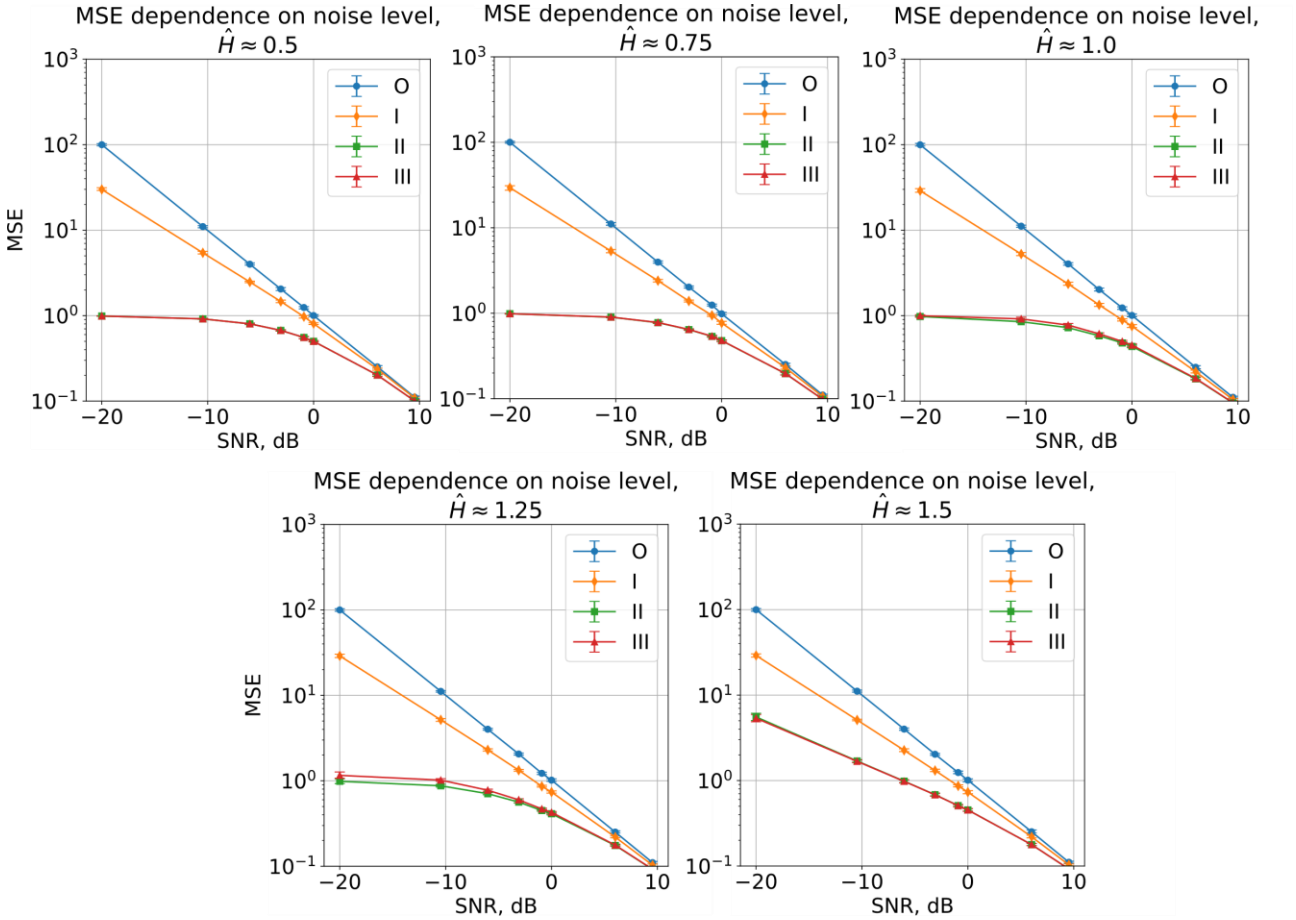

Figure 2. Comparison of the effectiveness of different signal recovery methods under varying noise levels and original signal parameters. For each value of the Hurst exponent $H = 0.5, 0.75, 1.0, 1.25, 1.5$, a signal is generated by the Kasdin model. Various noise levels are set from -20 to 10 dB. For each noise level, 15 noisy versions of the original signal are created and restored using three different Kalman filters (designated as I, II, and III according to the inherent model order). For each noise level, the error (MSE) between the original and noisy signals, the original and each of the three reconstructed signals is calculated with a 90% confidence interval (5th and 95th percentiles).

Figure 2 shows the dependence of the MSE (mean squared error) on SNR obtained for different H values with error bars denoting 5% and 95% quantiles, respectively. The figure shows that the first-order filter is unable to reconstruct the original signal, and thus the error increases with increasing disturbance. In contrast, second- and third-order filters are potentially capable of reconstructing the signal with accuracy bounded solely by the fluctuation error. A plateau at low SNR values could be attributed to the regime where measurements are given minimal weight, and thus the reconstruction quality is fully determined by the extrapolations. In marked contrast, for the non-stationary regime, corresponding to $H > 1$, in addition to the fluctuations there is an obvious accumulation of the regular bias, resulting in the linear increase of the MSE with increasing (equivalent) noise level.

Figure 3 shows how the MSE depends on the Hurst exponent H for fixed SNR levels and gap model parameters, respectively.

Figure 3 (*a*) summarizes the results obtained for a fixed noise level of -20 dB across different Hurst parameter (H) values. The figure shows that the MSE decreases with increasing filter order, while remaining independent of the Hurst exponent for all fractional Brownian motion scenarios, although increases at $H > 1$ due to the additional accumulation of the regular bias (see also Fig. 2).

Figure 3 (*b*) extends the previous noise analysis by examining how different Kalman filter models handle another common data problem – gaps and how their performance depends on the Hurst exponent. The gap occurrence frequency was set to 0.1, and the total number of test signals for each H is N = 10. The figure shows that the MSE reduces monotonously with both increasing the filter model order and the Hurst exponent. The latter could be attributed to the fact that large Hurst exponents correspond to strongly persistent data series with remarkably slower dynamics, in comparison with high-frequency white noise, and thus even simple linear extrapolations are performing rather well in the presence of gaps, reducing the effective error levels for Kalman filters of different orders.

In the following, we reconstruct the simulated signals using Kalman filters of the 1st, 2nd and 3rd orders. For each filter model, the MSE between the reconstructed and the original signals is calculated and visualized by boxplots.

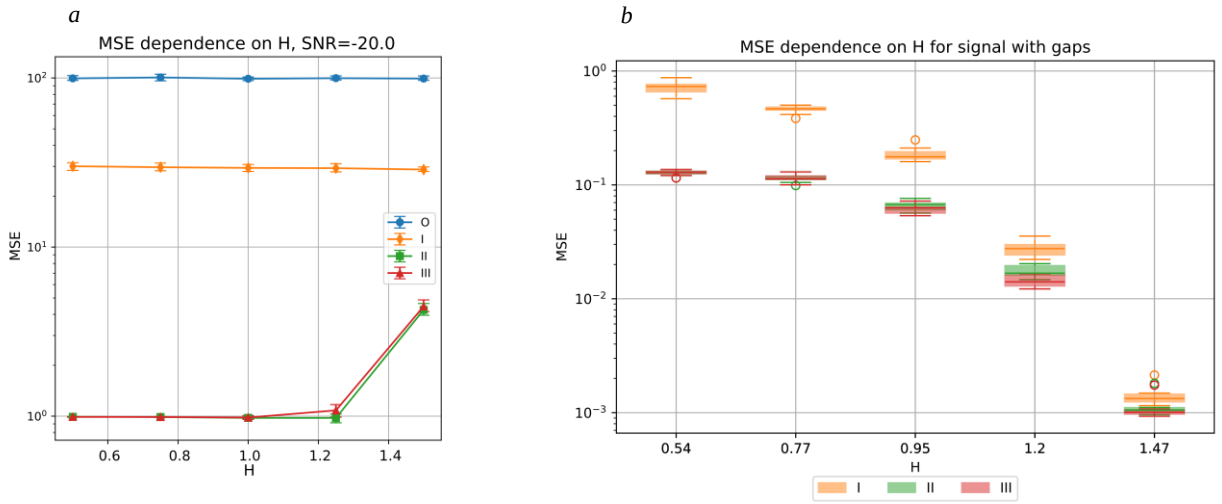

Figure 3. Dependences of the MSE on the value of the Hurst exponent with additive Gaussian noise and a signal-to-noise ratio of -20 dB (*a*) and with gaps (*b*)

5. Application to empirical video recordings

We use a simple algorithm that performs segmentation and quantification of heterogeneous regions in various biomedical images based on an estimate of the density of their local boundaries. The algorithm consists of several simple steps, including a conventional edge detector, followed by a smoothing filter that extracts local edge density values, which are eventually used for segmentation using a multi-threshold-based procedure.

To binarize the original monochrome image, a threshold T should be chosen. The image pixels that exceed the threshold will be set to 1 and will represent the objects of interest. The pixels that are less than the threshold will be set to zero and will correspond to the background of the image.

Too small threshold values result in the inclusion of background image fragments in the objects of interest, thus distorting them. Excessively large threshold values lead to disruption of the connectivity of the selected objects of interest and changes in their shape. Thus, an unjustified choice of threshold leads to an increase in the level of noise in the image after binarization and, accordingly, to a distortion of the selected objects of interest.

To reasonably choose the binarization threshold, we used a technique based on the application of the coefficient of elongation of the perimeter of the object (Bogachev, 2019; Volkov, 2020):

$$P_S = \frac{P^2}{4\pi S},$$

where P is the objects perimeter, S is its area. Of note, the above metric is also an inverse to the well-known circularity metric (Mendoza and Lu, 2015). This coefficient is normalized, and for objects that have a shape close to a circle, it tends to 1. When an object is lengthened or its boundaries are distorted, this coefficient increases. In practice, it is advisable to consider a series of values of the binarization threshold T, taken with a certain step ΔT. In our study, it is assumed that the objects of interest have a round shape. Thus the optimal threshold for image binarization will be one at which the values of the perimeter elongation coefficient for selected objects will be minimal.

Trajectory processing is the process of association of the specific objects selected in individual frames with their unique trajectories. It starts with the initialization of individual trajectories. We assume that, once there are samples in the processed frame that have not been identified with other trajectories, they should be considered the starting points of new trajectories.

After the initialization, selected objects in each incoming frame are associated with one of the existing trajectories. The simplest method that implements this stage is the Global Nearest Neighbor (GNN) method: the point closest to the last reference point of this trajectory will be considered a continuation of the trajectory. To increase the effectiveness of this method, we performed several actions: to search for the nearest point not to the last reference, but to the interpolated value based on the extrapolation by the proposed Kalman filter model, in turn calculated based on the previous object position and movement, as well as incoming measurements.

When no new points associated with a particular trajectory for a certain period of time could be detected, the trajectory is being terminated. However, a tracking failure situation is also possible – when the object continues to be in the field of view, but new readings for one or another reason do not appear due to errors in the computer vision system. To avoid this situation, one continues to interpolate the object's movement, and then, after a while, transfers the trajectory to the “lost” category – search for new points potentially belonging to this trajectory continues, while no further extension of the trajectory is being performed, until the receipt of any new observations that automatically transfer the trajectory back into the “observed” status.

Following the above strategy, an algorithm that processes the initial unstructured array of coordinates of potential objects into an array of trajectories of individual objects convenient for further analysis and processing has been implemented.

Figure 4 exemplifies the reconstruction of multiple animal movement trajectories based on the video capture of ants’ colony movements (*b*) from object selections in single frames shown in (*a*), respectively, indicating successfully resolved missing events, merged duplicates, as well as denoising based on the Kalman filter-based processing of trajectories. The

proposed approach is particularly suited for the analysis of long-term correlated motion patterns common in many natural and especially living systems, and thus provide a tailored approach to the trajectory analysis in biomedical and ecological research dealing with motion patterns of living objects from microscales (e.g., microbes or cells) to macroscales (e.g. animals in laboratory or natural living conditions).

a *b*

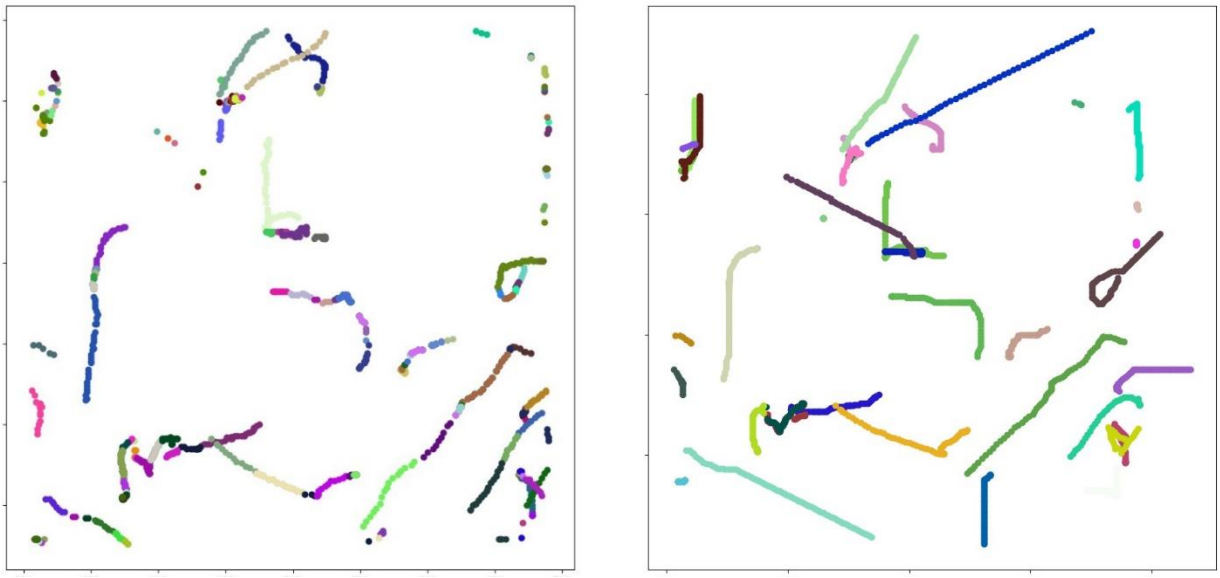

Figure 4. (a) Multiple object selections; (b) reconstructions of multiple animal movement trajectories.

Acknowledgements

We like to acknowledge the financial support of this work by the Ministry of Science and Higher Education of Russian under assignment FSEE-2025-0006.

References

Sinitca, A.M. et al. 2023a. Segmentation of patchy areas in biomedical images based on local edge density estimation. Biomedical Signal Processing and Control, 79, p.104189.

Sinitca, A. et al., 2023b, June. BCAnalyzer: A Semi-automated Tool for the Rapid Quantification of Cell Monolayer from Microscopic Images in Scratch Assay. In *International Work-Conference on Bioinformatics and Biomedical Engineering* (pp. 256-269). Cham: Springer Nature Switzerland.

Sinitca, A.M. et al. 2023c. Multi-Class Segmentation of Heterogeneous Areas in Biomedical and Environmental Images Based on the Assessment of Local Edge Density. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, pp.233-238.

Volkov, V.Y., Bogachev, M.I. and Kayumov, A.R., 2020. Object selection in computer vision: from multi-thresholding to percolation based scene representation. Computer Vision in Advanced Control Systems-5: Advanced Decisions in Technical and Medical Applications, pp.161-194.

Bogachev, M.I., Volkov, V.Y., Kolaev, G., Chernova, L., Vishnyakov, I. and Kayumov, A., 2019. Selection and quantification of objects in microscopic images: from multi-criteria to multi-threshold analysis. Bionanoscience, 9, pp.59-65.

F. Mendoza, R. Lu. Basics of Image Analysis. In: Hyperspectral Imaging Technology in Food and Agriculture. Ed. by Bosoon Park, Renfu Lu. Springer Science+Business Media New York 2015. DOI:10.1007/978-1-4939-2836-1_2

Bogachev, M.I. et al., 2023. Understanding the complex interplay of persistent and antipersistent regimes in animal movement trajectories as a prominent characteristic of their behavioral pattern profiles: Towards an automated and robust model based quantification of anxiety test data. Biomedical signal processing and control, 81, p.104409.

Kasdin, N.J., 1995. Discrete simulation of colored noise and stochastic processes and 1/f/sup/spl alpha//power law noise generation. Proceedings of the IEEE, 83(5), pp.802-827.

Jonker R., Volgenant T. A shortest augmenting path algorithm for dense and sparse linear assignment problems //DGOR/NSOR: Papers of the 16th Annual Meeting of DGOR in Cooperation with NSOR/Vorträge der 16. Jahrestagung der DGOR zusammen mit der NSOR. – Springer Berlin Heidelberg, 1988. – C. 622-622.