

Expectation maximization based framework for joint localization and parameter estimation in single particle tracking from segmented images - Simulation Data

Lin, Ye, Boston University, <https://orcid.org/0000-0003-2506-189X>

Andersson, Sean B., Boston University

yelin@bu.edu, sanderss@bu.edu

Publisher: Dryad

Journal

Plos One

Paper name

Expectation maximization based framework for joint localization and parameter estimation in single particle tracking from segmented images

Abstract of the datasets

The datasets contain simulated Single Particle Tracking (SPT) data consisting of sequences of camera images of a single fluorescent sub-diffraction limit-sized particle undergoing two-dimensional diffusion. We simulated a variety of experimental conditions, including different Signal-to-Background Ratios (SBRs), two different camera types, different diffusion speeds, and two different settings for motion blur. SPT is a class of experimental methods and data analysis techniques for exploring the motion of individual biological macromolecules. Typical estimation algorithms split the problem into two parts: first localize the particle at each data point to generate a trajectory and then estimate model parameters from that trajectory. We have recently introduced a class of algorithms for jointly estimating both trajectory and model parameters. In this study, we used the data to perform quantitative comparisons between two variants of our approach, one relying on a Sequential Monte Carlo methods combined with Expectation Maximization (SMC-EM) that is applicable to a very broad set of motion and observation models, and one that replaces the SMC elements with

methods based on the Unscented Kalman Filter (UKF) to improve upon the computational complexity. We also compared our methods to two current standards in the field. The first uses Gaussian Fitting to localize the particle, following by a Mean Square Displacement (GF-MSD) analysis to determine model parameters while the other replaces MSD with Maximum Likelihood Estimation (GF-MLE). The main results of our study indicate that our EM-based schemes significantly outperform the existing algorithms at low SBR while at high SBR, GF-MLE performs equally well but at a lower computational cost.

Methods

All datasets were simulated using MATLAB (MathWorks, Natick, MA). Two camera types were considered: (1) An ideal camera with Poisson distributed shot noise but no readout noise; (2) A camera with both shot noise and the pixel-dependent readout noise that is common to scientific complementary metal-oxide semiconductor (sCMOS) cameras. For each run, the diffusive motion of a single particle was generated at a time step of 1 ms. At each time step, a pixelated image was generated by integrating a standard Gaussian model of the Point Spread Function (PSF) of the microscope over each of the pixels in the camera image. The cameras were assumed to take images at a rate of 10 Hz with a shutter period of 10 ms. For data that included motion blur, the 10 pixelated images in the shutter period were accumulated to generate a single camera image. For data that ignored motion blur, only the first pixelated image was used. In each case, both background and camera readout noise were included based on the model of the camera being considered. Datasets were generated at a range of SBRs and diffusion coefficients (see Usage Notes for details).

Usage Notes

* DATA-SPECIFIC INFORMATION FOR: [type1_data_with_different_N_G.zip]

1. Images were captured by the camera with Poisson distributed shot noise but no readout noise.
2. The background noise ranges from 1 to 15, while the signal level ranges from 1 to 100.
3. Motion blur is considered (each image contains 100 sub-samples, and only the first 10 sub-samples of every 100 sub-samples were accumulated).

4. Taking the sub-directory [N10G100Image100] as an example, the dataset name indicates the following parameter values: background noise=10, signal intensity=100, number of images per dataset=100. Under each experimental setting, 100 datasets were simulated.
5. Taking the first dataset under [N10G100Image100] as an example, the details of the provided data are as follows:
 - [photon_observation_1.csv] row*column=100*25; There are the simulated images with each row containing a single image, organized sequentially from the initial time down to the final time. The 25 entries in each row are the intensity values from the 5-by-5 pixel array, organized starting from the top-left pixel, going horizontally across the top five pixels, and continuing to the bottom right pixel.
 - [sensor_position_1.csv] row*column=100*2; In generating images, we assumed segmentation of the full camera image was done previously and thus the location of the 5-by-5 array of pixels may change at each time step. This data gives the position of the pixel array starting from the initial time down to the final time, with the first column corresponding to the x-coordinate and the second column to the y-coordinate.
 - [x_ground_truth.mat] row*column=100*100; This data contains the true location of the particle in the x-direction at each time step. Because of the motion blur model, this location was taken to be the mean of the position across the shutter period. The data is organized starting with the first time step and proceeding down to the last.
 - [y_ground_truth.mat] row*column=100*100; This data is the true location of the particle in the y-direction at each time step, and it is organized as [x_ground_truth.mat].

* DATA-SPECIFIC INFORMATION FOR:[type1_N10G100Image100D_different.zip]

1. Images were captured by the camera with Poisson distributed shot noise but no readout noise.
2. Background noise=10, signal intensity=100, number of images per dataset=100, diffusion coefficient ranges from 0.001 to 10.

3. The sub-directory [100datasets_withSub] contains the data with motion blur, while the sub-directory [100datasets_noSub] contains the data without motion blur.
4. Taking the files in [100datasets_withSub/N10G100D0.01] as an example, the diffusion coefficient is $D_x=D_y=0.01(\mu\text{m}^2/\text{s})$. The variable names and their corresponding dimensions are the same as the descriptions of [type1_data_with_different_N_G.zip].

* DATA-SPECIFIC INFORMATION FOR: [type2_camera.zip]

1. Images were captured by the camera with both shot noise and the pixel-dependent readout noise.
2. [N10G100Image100sCMOS] background noise=10, signal intensity=100, number of images per dataset=100. For each experimental setting, 100 datasets were simulated. Taking the 1st dataset as an example, the provided data is:
 - [local_sigma_1.csv] row*column=100*25; row denotes the time steps, column denotes the pixel. Each pixel contains information of "sigma" characterizing the readout noise of the Hamamatsu ORCA Flash 4.0 camera.
 - [photon_observation_1.csv] are the measured intensities, organized as described above.
 - [sensor_position_1.csv] are the global locations of the 5-by-5 pixel arrays, organized as described above.
 - [x_ground_truth.mat] are the ground truth in the x-direction, organized as described above.
3. [N10G30Image100sCMOS] background noise=10, signal intensity=30, number of images per dataset=100. For each experimental setting, 100 datasets were simulated. This data is organized in the same manner as described above.

* DATA-SPECIFIC INFORMATION FOR: [Data_for_figures_PlosOne.zip]

Figures and their corresponding data used for reproducing the figures in paper "Ye Lin and Sean B. Andersson. Expectation maximization based

framework for joint localization and parameter estimation in single particle tracking from segmented images. Plos One (2021).”

Data Citation

Lin, Ye; Andersson, Sean B. (2021), Expectation maximization based framework for joint localization and parameter estimation in single particle tracking from segmented images - Simulation Data, Dryad, Dataset, <https://doi.org/10.5061/dryad.9w0vt4bf5>

Funding

National Institutes of Health, Award: 1R01GM117039-01A1