ACCURATE ESTIMATION OF PARTICLE DYNAMICS BYPASSING SUBSTRATE DRIFT BIAS: APPLICATION TO CELL NUCLEUS MOTION

Hélène Kabbech, Jente van Staalduinen, Frank Grosveld and Ihor Smal

Department of Cell Biology, Erasmus University Medical Center, Rotterdam, the Netherlands

ABSTRACT

In microscopic imaging, the movement of a living substrate can be caused by its own displacement (e.g., cell motion/migration) or other technical factors such as microscope stage drift. This drifting motion is one of the main biases resulting in poor estimation of particle dynamics since it seriously affects the estimation of the biophysical parameters (the diffusion constant D and anomalous exponent α), especially when performed on the basis of mean squared displacement (MSD) analysis. In this paper, we compare a few substrate drift correction/registration methods based on the use of additional fluorescent spots (landmarks). In the particular case of cell nucleus motion, we labeled telomeres spreading throughout the cell nucleus. We show that compared to the MSD analysis, the use of Gaussian processes is an effective and more accurate way to estimate the substrate drift, and major biophysical parameters of particle dynamics.

Index Terms— Single-particle tracking, registration, mean squared displacement, Gaussian processes, cell nuclei

1 Introduction

Image registration is the process of transforming the coordinates of an image (the *moving* image) to align with a second one (the *fixed* image) by applying a series of geometric transformations or local displacements. This problem occurs frequently in many fields, including biomedical imaging (e.g., registration of CT/MRI images of brain). Commonly used geometric transformations include *rigid* body motion, where two images are transformed using only rotation and translation, or *non-rigid*, where any local transformation is allowed, defined by the "field" of deformations between the two images. In addition, registration methods can be *intensity-based* strictly using the intensity patterns in the images, or *feature-based* using features such as landmarks, lines and points.

Registering the motion of a living substrate (e.g., cell nuclei) imaged with fluorescence microscopy is known to be difficult [1]. The registration of a stack of images is more complex to achieve in terms of assembling the transformed images, also requiring more computation time with intensity-based methods. The latter require the presence of characteristic intensity patterns (texture) that are typically missing in images of cell. Nonetheless, depending on the staining procedure, it is possible to label at least some of the regions, for example the contour of the nuclear envelop based on a fluorescent DNA dye [2]. Yet, forces driven by both chromatin and cytoskeleton cause the shape of the nuclear envelope to dynamically fluctuate over time [3], complicating the possibility to use rigid body image-based registration.

Since no fixed characteristic structure is present in live cells as a reference for matching alignment, more recent methods use additional fluorescently labeled spots (as markers) in the nucleus, for example, labeled chromatin loci [4, 5]. In order to estimate the global drift of the nucleus those "landmarks" are used for image transformations in combination with so-called point cloud registration methods, which use rigid/affine transformations [5] or Gaussian Processes (GPs) [4]. Here, we propose to use labeled chromosome-ends (telomeres) as markers which can be used by point cloud registration methods to estimate substrate drift and accurately estimate motion parameters. Our staining procedure provides a good amount of (diffraction-limited) spots localized throughout the nuclear region and stable over time with a relatively high signal-to-noise ratio. We compare a few registration methods proposed in the literature (averaging the tracks, applying affine transformations and Gaussian processes), using both synthetic data and telomere trajectories from real experiments, and compare the estimated motion parameters before and after applying each method. We conclude that registration of the substrate drift is more than necessary for an accurate estimation of the α parameter using the MSD analysis. The use of Gaussian processes is an effective way to estimate smoothed drifts. It can also be used to accurately estimate the motion parameters without prior registration.

2 Methods

2.1 Problem statement

Particle dynamics can be characterized by the diffusion constant D and anomalous exponent α, which can be both estimated using a least-square fit to the logarithm form of the mean squared displacement: $\mathrm{MSD}(\Delta t) \sim D\Delta t^{\alpha}$, where $0 < \alpha < 2$. The position of diffraction-limited fluorescent particles in each image frame at time t can be accurately estimated by fitting a Gaussian Point Spread Function (PSF) and then linked to form trajectories using the nearest-neighbor crite-

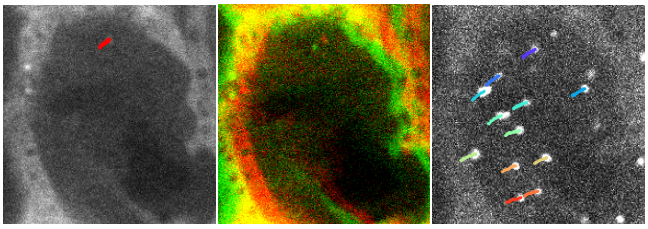

Fig. 1: Particles diffusing within a cell nucleus for 75 s and manifesting a strong drift caused by the cell motion. (left) Trajectory of a single chromatin locus. (middle) Superposition of the first (green) and last (red) frames to reveal the drift. (right) Trajectories of telomere spots.

rion [6]. The result of the particle tracking is a set $\mathcal{S}$ of trajectories $s_i \in \mathcal{S}$, $i = \{1, \ldots, P\}$, where P is the number of particles, and $s_i(t) = (x_i(t), y_i(t))$ represents the coordinates of a particle at time t. Each particle s_i diffuses in a cell nucleus, which adds an unknown global motion (translation, rotation) to all the trajectories $\mathcal{S}$ caused by the cell motion and any other technical drifts (see example in Fig. 1). For instance, in the case of a strong drift, the particles might be wrongly classified as undergoing *superdiffusion*. Correcting for such global motion corresponds to finding geometric transformations which can be used to "subtract" this bias from the particle trajectories, and thus end up with accurate local behavior. Such transformation can be achieved in several ways, discussed below.

2.2 Trajectory averaging

The simplest registration method, which was previously applied to telomere trajectories [7], consists of averaging all particle positions in every frame. The averages $(m_x(t), m_y(t)) = (\frac{1}{P}\sum_{n=1}^{P} x_i(t), \frac{1}{P}\sum_{n=1}^{P} y_i(t))$ represent the estimated global motion, which is later subtracted from the initial trajectories in $\mathcal{S}$.

2.3 Affine transformations

An affine transformation is a geometric linear mapping that preserves parallelism and ratios of distances between points, lines and planes. Additionally to the rigid transformation, which includes only rotation and translations, the affine transformation can also handle scaling and shearing. Finding the transformations between two sets of points from two frames can be done very efficiently, resulting in estimated global motion on a per-frame basis [5].

2.4 Gaussian processes

A Gaussian process (GP) is a distribution over stochastic functions $f(t) \sim \mathcal{GP}(m(t), k(t, t'))$ defined by its mean function $m(t)$ and covariance function (kernel) $k(t, t')$ [8]. A GP can be used for different estimation purposes depending on the chosen mean function and kernel. In our case we used GPs to estimate the drift for each trajectory. For this purpose, we used a standard GP model with a zero mean function and the radial basis function (RBF) kernel, $k_{\text{RBF}}(t, t') = \exp(-||t - t'||^2/2l^2)$ where l is the lengthscale parameter which represents the smoothness of the estimated function that models the global drift. We used a lengthscale value of 100 and fed the GP with trajectory displacements. After estimating the global drift for each track (refered as GP-RBF), we either subtract the estimates from the corresponding trajectories, or use them as inputs to estimate the affine transformations.

Within the GP framework, it is also possible to use the kernel which describes the fractional Brownian motion (FBM), the process that is frequently used in practice to model anomalous diffusion [9]. For this case, the kernel is defined as $k_{\text{FBM}}(t, t') \sim D\left[|r+1|^{\alpha} - 2|r|^{\alpha} + |r-1|^{\alpha}\right]$ with $r = |t - t'|$. This model (referred as GP-FBM) can be used to directly estimate D and α from the trajectories that contain a global drift [4]. For each coordinate, $x_i(t)$ and $y_i(t)$, we built a GP model with a $m(t) = 0$ and the FBM kernel, and fed the corresponding displacements. The optimal values for the motion parameters can be estimated by maximizing the log marginal likelihood. The estimated parameters for x- and y-dimensions were averaged. We compared our GP-FBM model with the software package GP-Tool [4], which also uses GPs but in a different setup and works only with 2-5 trajectories per image sequence.

3 Results

A set of 5 trajectories (300 frames long) was generated following the FBM diffusion model [9] with $\alpha = 0.5$ and $D = 0.045$ [pixel^2/frame]. A global motion was simulated by applying a constant translation and rotation at each frame. The total translations in the x- and y-dimensions are 20 and -20 pixels, while the total rotation is 15° (Fig. 2a, top panel).

Telomeres in mouse embryonic stem cells (mESC) were labeled by PiggyBac-mediated stable expression of a hCdt(1-100)-EGFP-mTERF2 expression construct and imaged on a Nikon Eclipse Ti-E inverted microscope with a Yokogawa dual spinning-disc confocal scanning unit (Yokogawa, CSU-X1-A1) and a Nikon CFI APO TIRF 100x 1.49 NA oil objective at 37°C and 5% CO2. Cells were excited by a 491 and 561 nm diode laser and images were captured simultaneously with a DV2 Beamsplitter (MAG Biosystems) on a QuantEM512C camera (Photometrics). Stream acquisitions were acquired in two dimensions for 75 s, with a time interval of 250 ms (300 frames in total). The chromatin locus was labeled using the ANCHOR3 DNA labeling system [10]. The tracking was done using the SOS tracker [6] (Fig. 1).

3.1 Synthetic data

We applied the registration techniques described above to the simulated trajectories (the estimated drifts and registered tracks are shown in Fig. 2b-e), and computed the motion parameters of interest using the MSD-based approach. The absolute errors in estimating α and D are shown in Fig. 3

(left panels). Additionally, we estimated the parameters for the raw trajectories before and after adding the drift, using GP-FBM [4] and our own implementation (Fig. 3, right panels).

The averaging method produces a single drift estimate, and is therefore unable to correct the drift in the presence of a substrate rotation (Fig 2b). The estimates generated with affine transformation are better but rather noisy (Fig 2c). The GP-RBF method estimated drifts using smooth curves (Fig 2d). Combining the GP-RBF estimates with the affine transformations helps in keeping both the smoothness of drifts and the correct pathways (Fig 2e).

Compared to the MSD method, our GP-FBM produces very accurate estimates for α with and without the additional drifts. At the same time, for both the MSD and our GP-FBM, the estimates of D are quite comparable, with and without the drift. The estimates given by the GP-FBM [4] are rather inaccurate for both α and D on our set of trajectories.

3.2 Real data

We also applied the described methods to the analysis of telomere dynamics, where the ground-truth information about the global motion and other parameters are unknown. In principle, the underlying α and D might vary per telomere spot.

For the time lapse shown in Fig. 1, the results of drift estimation are given in Fig. 4, and the parameter estimates for the registered telomeres (tel 1-5) are given in Table 1. The α estimates are in the same range for I to IV-MSD and our GP-FBM implementation. Similarly, the D estimates are in the same range for the MSD, our GP-FBM, III-MSD and IV-MSD. The GP-FBM from Oliveira et al (2021) produced different ranges of α and D estimates.

We also applied the transformations to our spot of interest (a single chromatin locus, Fig. 1a) and estimated its dynamics (Table 1, loc), resulting in α of about 0.1 and a diffusion constant D of 0.0029 $\mu m^2/s$, which are interestingly rather different from the dynamics of the telomere spots.

	Track id	tel 1	tel 2	tel 3	tel 4	tel 5	loc
α_{pred}	MSD	0.475	0.465	0.327	0.511	0.598	0.178
	GP-FBM [4]	0.357	0.285	0.324	0.363	0.569	—
	GP-FBM (ours)	0.264	0.178	0.188	0.279	0.416	0.091
	I-MSD	0.297	0.198	0.149	0.245	0.326	0.112
	II-MSD	0.299	0.157	0.154	0.258	0.290	0.127
	III-MSD	0.262	0.130	0.135	0.234	0.385	0.098
	IV-MSD	0.286	0.140	0.138	0.264	0.395	0.114
D_{pred}	MSD	5.7	3.8	10.2	3.4	4.1	27.8
	GP-FBM [4]	4.8	3.1	5.6	2.2	3.1	—
	GP-FBM (ours)	6.1	4.5	10.6	3.6	4.3	28.7
	I-MSD	4.3	3.1	7.0	3.1	3.4	29.3
	II-MSD	4.5	2.0	3.8	3.2	2.7	31.6
	III-MSD	5.9	4.0	10.6	3.6	4.1	29.1
	IV-MSD	5.9	3.8	10.6	3.4	4.1	28.2

Table 1: Estimated motion parameters on 5 telomere trajectories (tel) and our chromatin locus of interest (loc). D_{pred} values are in $[10^{-4} \times \mu m^2/s]$.

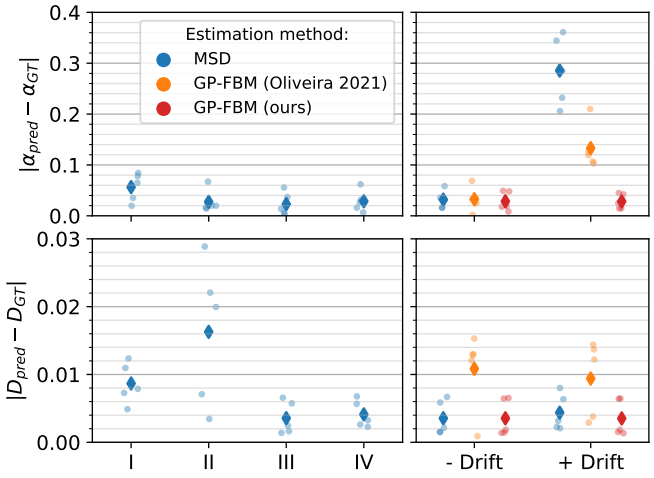

Fig. 3: Absolute errors in the α (top) and D (bottom) estimates for 5 simulated trajectories. (left panels) Registered tracks per point cloud registration methods: (I) AVG, (II) AT, (III) GP-RBF, (IV) GP-RBF + AT. (right panels) Raw tracks with or without the additional substrate drift.

4 Discussion and Conclusion

Several conclusions can be drawn from this comparative study: (i) the MSD analysis, which is the most commonly used method for estimation of α and D, cannot cope with a global drift (for example due to substrate motion), and will overestimate α; (ii) labeling the telomeres is an efficient way to obtain multiple trajectories in live cells and successfully correct for nuclear drift using point cloud registration methods. This allowed us to estimate the dynamics of a chromatin locus after correction of substrate drift; (iii) the averaging of the tracks does not provide accurate registration if the substrate is rotating. The affine/rigid transformations produce noisy drift estimates due to the independent calculation for each pair of frames. The GP-RBF produces smoothed drift estimates, which in combination with the affine transformation could help in eliminating the local uncertainty bias of the affine transforms; (iv) finally, directly estimating the motion parameters with our GP-FBM implementation gave very accurate estimates in all considered cases. Compared to Oliveira et al (2021), our implementation can work for a single track and is not dependent on other trajectories located within the same substrate. Estimation of the global motion can still be of interest for other purposes, for instance to analyze the movement of the substrate itself over longer periods of time.

Acknowledgments. This work was supported by the Dutch Research Council through the NWO-BBOL research program (Grant No. 737.016.014).

Compliance with Ethical Standards. This is a numerical simulation study for which no ethical approval was required.

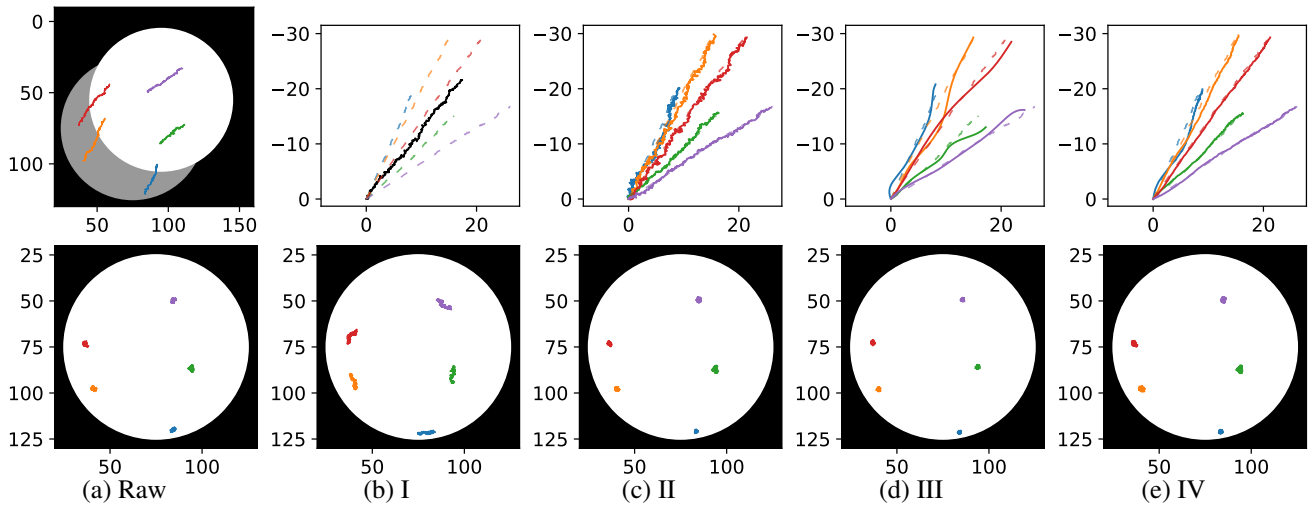

Fig. 2: (a) Simulated tracks with (top) and without (bottom) an additional nucleus drift. (b-e) Estimated drift for each track with the ground-truth drifts in dashed line (top), registered tracks (bottom) using one of the following registration technique: track averaging [AVG] (b); affine transformations [AT] (c); Gaussian process estimates with RBF kernel [GP-RBF] (d); affine transformations using the Gaussian process estimates as inputs [GP-RBF + AT] (e).

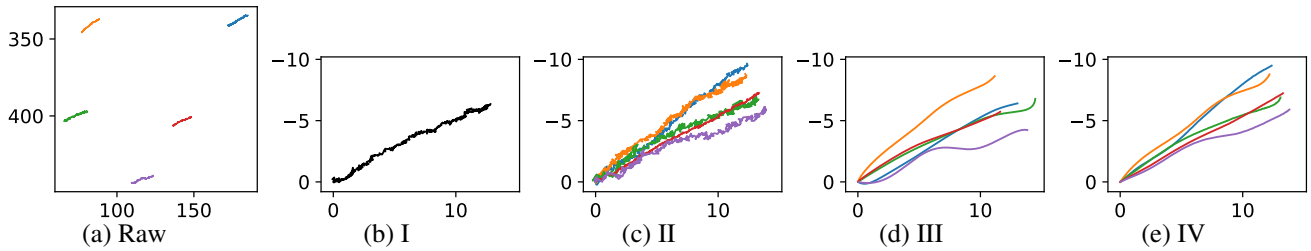

Fig. 4: (a) Telomere tracks. (b-e) Estimated drifts using one of the point cloud registration technique (see caption in Fig. 2).

5 References

[1] S. Yang et al., "Nonrigid registration of 3-d multichannel microscopy images of cell nuclei," *IEEE Trans. on Im. Proc.*, vol. 17, pp. 493–499, 2008.

[2] D.V. Sorokin et al., "Non-rigid contour-based registration of cell nuclei in 2-D live cell microscopy images using a dynamic elasticity model," *IEEE Trans. Med. Im.*, vol. 37, pp. 173–184, 2017.

[3] F-Y. Chu et al., "On the origin of shape fluctuations of the cell nucleus," *Proc. of Nat. Acad. of Sci.*, vol. 114, pp. 10338–10343, 2017.

[4] G.M. Oliveira et al., "Precise measurements of chromatin diffusion dynamics by modeling using gaussian processes," *Nat. Com.*, vol. 12, pp. 1–11, 2021.

[5] P. Mach et al., "Cohesin and ctcf control the dynamics of chromosome folding," *Nat. Genet.*, pp. 1–12, 2022.

[6] E. Meijering et al., "Methods for cell and particle tracking," *Methods in Enzym.*, vol. 504, pp. 183–200, 2012.

[7] L. Stadler and M. Weiss, "Non-equilibrium forces drive the anomalous diffusion of telomeres in the nucleus of mammalian cells," *New J. of Phys.*, vol. 19, pp. 113048, 2017.

[8] I. Smal et al., "Gaussian processes for trajectory analysis in microtubule tracking applications," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 206–209.

[9] T. Lundahl et al., "Fractional brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. on Med. Im.*, vol. 5, pp. 152–161, 1986.

[10] T. Germier et al., "Real-time imaging of specific genomic loci in eukaryotic cells using the anchor dna labelling system," *Methods*, vol. 142, pp. 16–23, 2018.