Rao-Blackwellized Marginal Particle Filtering for Multiple Object Tracking in Molecular Bioimaging

Ihor Smal[1], Katharina Draegestein[2], Niels Galjart[2],
Wiro Niessen[1], Erik Meijering[1]

[1]Departments of Radiology and Medical Informatics
[2]Department of Cell Biology
Erasmus MC – University Medical Center Rotterdam
P. O. Box 2040, 3000 CA Rotterdam, The Netherlands
Correspondence: i.smal@erasmusmc.nl

Abstract. Modern live cell fluorescence microscopy imaging systems, used abundantly for studying intra-cellular processes in vivo, generate vast amounts of noisy image data that cannot be processed efficiently and accurately by means of manual or current computerized techniques. We propose an improved tracking method, built within a Bayesian probabilistic framework, which better exploits temporal information and prior knowledge. Experiments on simulated and real fluorescence microscopy image data acquired for microtubule dynamics studies show that the technique is more robust to noise, photobleaching, and object interaction than common tracking methods and yields results that are in good agreement with expert cell biologists.

Key words: Bayesian estimation, particle filtering, multiple object tracking, Rao-Blackwellization, microtubule dynamics, fluorescence microscopy.

1 Introduction

Live cell imaging using time-lapse fluorescence microscopy has rapidly advanced in the past decade and has opened new possibilities for studying intra-cellular dynamic processes in vivo. Motion analysis of nanoscale objects, such as proteins, vesicles, or microtubules (Fig. 1), requires tracking of large and time-varying numbers of spots in noisy image sequences [1–3]. Manual analysis of such image data is laborious and often produces results with poor accuracy and/or reproducibility. Hence, the development of automated tracking methods is of great importance. Commonly used tracking methods fail to yield reliable results in the case of poor imaging conditions (SNR<5) [4], because the detection is usually based on simple intensity thresholding or model fitting, and available temporal information and prior knowledge are largely ignored. Alternative techniques, based on spatiotemporal segmentation [5], are also prone to errors in the case of very noisy images containing many objects at high densities.

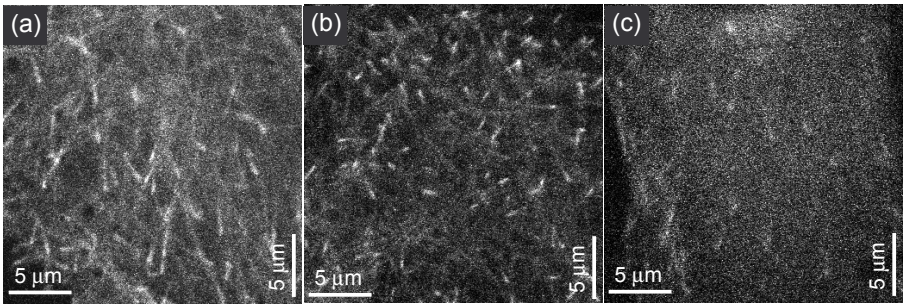

Fig. 1. Microtubules tagged with fluorescently labeled plus end tracking proteins (bright spots) and imaged using fluorescence confocal microscopy. The images are single frames from three 2D time-lapse studies, acquired under different experimental conditions. The quality of the images ranges from SNR≈5–6 (a) to ≈2–3 (c).

Recently, sequential Monte Carlo (SMC) methods [6], also known as particle filters (PF) [7], have become a popular tool to perform tracking in many fields. In this paper we extend our previous PF approach [8] and present a substantially more efficient Rao-Blackwellized marginal particle filter (RBMPF) for robust and accurate tracking of multiple nanoscale targets in two-dimensional (2D) and three-dimensional (3D) fluorescence microscopy image sequences. The RBMPF takes into account the analytical structure of the modeled processes and makes it possible to reduce the variance of the estimates in the case of high-dimensional state spaces, where standard PF fails. We compare the performance of standard PF and RBMPF with manual tracking using simulated as well as real image data acquired for microtubule dynamics studies.

2 Tracking Framework

Bayesian estimation for tracking aims at inferring knowledge about the unobserved state $\mathbf{x}_t$ of an object, which changes over time, using noisy measurements $\mathbf{z}_{1:t} \triangleq \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ up to time t. The state evolution is modeled as a Markov process of initial distribution $p(\mathbf{x}_0)$ and transition prior $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. The idea is to sequentially estimate the time evolving joint filtering distribution $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$ or the marginal filtering distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ and associated features, such as expectation. A recursive formula for the former is given by [6]

$$p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1}|\mathbf{z}_{1:t-1}). \tag{1}$$

It is assumed that the initial pdf, $p(\mathbf{x}_0|\mathbf{z}_0) \equiv p(\mathbf{x}_0)$ is available ($\mathbf{z}_{1:0} = \mathbf{z}_0$ being the set of no measurements). The distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ follows from (1) as

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1}. \tag{2}$$

The optimal Bayesian solutions defined by the recurrence relations (1) and (2) are analytically tractable only in a restrictive set of cases [7]. For most practical models of interest, SMC methods [7,9] are used as an efficient numerical approximation. Here, the required posterior, $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$, is represented as a set of N_s random samples (particles), and associated weights $\{\mathbf{x}_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^{N_s}$:

$$p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) \approx \sum_{i=1}^{N_s} w_t^{(i)} \delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(i)}), \tag{3}$$

where $\delta(\cdot)$ is the Dirac delta function and the weights are normalized such that $\sum_{i=1}^{N_s} w_t^{(i)} = 1$. These samples and weights are propagated through time to give an approximation of the filtering distribution at subsequent time steps. The weights $w_t^{(i)}$ are chosen using sequential importance sampling (SIS) [10], which applies when auxiliary knowledge is available in the form of an importance density $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{z}_{1:t})$ describing which areas of the state-space contain most information about the posterior. In order to calculate the weights recursively, the importance density is factorized as $q(\mathbf{x}_{0:t}|\mathbf{z}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{z}_{1:t}) q(\mathbf{x}_{0:t-1}|\mathbf{z}_{1:t-1})$. The particle representation of the posterior at time t is obtained by augmenting the set of existing particles $\mathbf{x}_{0:t-1}^{(i)}$ with the new state $\mathbf{x}_t^{(i)} \sim q(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{z}_{1:t})$. A detailed formulation of $q(\cdot|\cdot)$ is given in Section 3.4. The weights $w_t^{(i)}$ in (3) may be recursively updated as [10]

$$w_t^{(i)} \propto \frac{p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})} = \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{z}_{1:t})} w_{t-1}^{(i)}. \tag{4}$$

In order to obtain the particle representation of $p(\mathbf{x}_t|\mathbf{z}_{1:t})$, only $\mathbf{x}_t^{(i)}$ need to be stored and the *path* $\mathbf{x}_{0:t-1}^{(i)}$ can be discarded. Each particle at time t, used to augment the previous state, is a draw from the joint space $p(\mathbf{x}_{0:t}|\mathbf{z}_{1:t})$, sampled sequentially. At each time step, the dimension of the sampled paths is increased by the dimension of the state space, n_x, quickly resulting in a very high-dimensional space. Because of the sequential nature of the algorithm, the variance of the importance weights can only increase (stochastically) over time [10], leading to most paths having vanishingly small probability. The degeneracy effect can be reduced by a good choice of importance density (Section 3.4) and by resampling [7, 10] to eliminate particles with small weights. Additionally, it can be reduced by using the marginal particle filter (MPF) [11], where the filtering is performed directly on the marginal distribution $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ defined by (2) instead of on the joint state. Having a representation of $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ in the form of (3), we can approximate the integral in (2) as the weighted kernel estimate $\sum_{j=1}^{N_s} w_{t-1}^{(j)} p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(j)})$. The importance weights are now on the marginal space:

$$w_t^{(i)} \propto \frac{p(\mathbf{x}_t|\mathbf{z}_{1:t})}{q(\mathbf{x}_t|\mathbf{z}_{1:t})} = \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(i)}) \sum_{j=1}^{N_s} w_{t-1}^{(j)} p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(j)})}{q(\mathbf{x}_t^{(i)}|\mathbf{z}_{1:t})}. \tag{5}$$

The variance of the importance weights is less than or equal to the variance of the standard SIS scheme [11]. By using particle representations, statistical inferences

such as expectation, maximum a posteriori, and minimum mean square error (MMSE) estimators, can easily be approximated. For example,

$$\hat{\mathbf{x}}_t^{\mathrm{MMSE}} = \int \mathbf{x}_t p(\mathbf{x}_t|\mathbf{z}_{1:t}) d\mathbf{x}_t \approx \sum_{i=1}^{N_s} \mathbf{x}_t^{(i)} w_t^{(i)}. \tag{6}$$

Straightforward generalization of SMC methods to the problem of multiple object tracking leads to an increase in dimensionality and an exponential explosion of computational demands [6]. We represent the multi-modal posterior distribution, in which multiple modes are caused by ambiguity about the object state due to either insufficient measurements or measurements coming from multiple objects being tracked, by an M-component mixture model [12]:

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \sum_{m=1}^{M} \pi_{m,t} p_m(\mathbf{x}_t|\mathbf{z}_{1:t}), \tag{7}$$

with $\sum_{m=1}^{M} \pi_{m,t} = 1$ and a non-parametric model is assumed for the individual mixture components. In this case, the particle representation of the filtering distribution, $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^{N}$ with $N = MN_s$ particles, is augmented with a set of component indicators, $\{c_t^{(i)}\}_{i=1}^{N}$, with $c_t^{(i)} = m$ if particle i belongs to mixture component m. This representation can be updated in the same fashion as the standard Bayesian sequential estimation [12].

3 Tailoring the Framework

3.1 State-Space and Dynamics

In this paper the framework is tailored towards microtubule (MT) tracking. MTs are cylindrical structures (diameter ∼25 nm) in the cytoskeleton that play a crucial role in several cellular processes [13]. We approximate the dynamic behavior of the visible ends of MTs by a nearly constant velocity model [14] with the state vector $\mathbf{x}_t = (x_t, \dot{x}_t, y_t, \dot{y}_t, z_t, \dot{z}_t, \sigma_{\max,t}, \sigma_{\min,t}, \sigma_{z,t}, I_t)^T$, where $(\sigma_{\max,t}, \sigma_{\min,t}, \sigma_{z,t})^T \triangleq \mathbf{s}_t$ is the object shape feature vector (Section 3.2), $(x_t, y_t, z_t)^T \triangleq \mathbf{r}_t$ is the radius vector, $\dot{\mathbf{r}}_t \triangleq \mathbf{v}_t$ velocity, and I_t object intensity.

In practice, the analysis of time-lapse fluorescence microscopy image sequences is complicated by photobleaching, a light-induced chemical process by which fluorescent proteins lose their ability to fluoresce. The two commonly used approximations of the process are given by [15]

$$I(t) = Ae^{-at} + B, \qquad \text{and} \qquad I(t) = I_0 \left(1 + (t/L)^k\right)^{-1} \tag{8}$$

where A, B, a, I_0, L, k are experimentally determined constants. Photobleaching is ignored by common tracking techniques but in many practical cases it is necessary to model it so as to increase robustness. To conveniently incorporate the photobleaching effect into our framework, we approximate it as a first-order Gauss-Markov process, $I_t = (1 - \alpha)I_{t-1} + u_t$, where u_t is zero-mean Gaussian noise and $\alpha \leq 1$ an experimentally obtained rate of photobleaching, which can be estimated from image data, using (8), by model fitting.

In summary, in our implementation, the state evolution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is a linear Gaussian model [10], which can easily be evaluated pointwise in (4):

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}_t - \mathbf{F}\mathbf{x}_{t-1})^T\mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{F}\mathbf{x}_{t-1})\right), \tag{9}$$

with process transition matrix $\mathbf{F} = \mathrm{diag}[\mathbf{F}_1, \mathbf{F}_1, \mathbf{F}_1, 1, 1, 1, 1-\alpha]$ and covariance matrix $\mathbf{Q} = \mathrm{diag}[\mathbf{Q}_1, \mathbf{Q}_1, \mathbf{Q}_1, q_2T, q_2T, q_2T, q_3T]$. Here,

$$\mathbf{F}_1 = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{Q}_1 = \begin{pmatrix} \frac{q_1}{3}T^3 & \frac{q_1}{2}T^2 \\ \frac{q_1}{2}T^2 & q_1T \end{pmatrix},$$

where q_1, q_2 and q_3 denote the level of process noise in object motion, appearance and intensity, respectively, and T is the sampling interval. The proposed model correctly approximates small accelerations in object motion and fluctuations in object appearance and intensity. To obtain a more realistic motion model and avoid track coalescence in the case of multiple objects, we explicitly model object interaction using a Markov random field (MRF) [8, 16]. The model (9) can also be successfully used for tracking structures with higher motion nonlinearity, by adapting the process noise level, defined by $\mathbf{Q}$.

3.2 Observation Model

The image formation process can be modeled as a convolution of the true, unobserved 3D image with the point-spread function (PSF) of the microscope. The PF framework can accommodate any PSF that can be calculated pointwise. Despite its minor imperfection, the 3D Gaussian approximation of the PSF [3] is commonly favored over the more accurate Gibson-Lanni model [1], for its computational advantages. To model the shape of the intensity profile of an imaged object, one would have to use the convolution with the PSF for every state $\mathbf{x}_t^{(i)}$. To overcome this computational overload, we model the PSF together with object shape using a 3D Gaussian approximation. The elongation in the intensity profile of MTs can be modeled by utilizing the velocity components from $\mathbf{x}_t$ as parameters in the PSF. In this case, for an object of intensity I_t at position $\mathbf{r}_t$, the intensity contribution to pixel (i, j, k) is approximated as

$$h_t(i, j, k; \mathbf{x}_t) = a_t(i, j, k; \mathbf{r}_t, \mathbf{v}_t, \mathbf{s}_t)I_t + b_t, \tag{10}$$

where b_t is the background intensity and

$$a_t(i, j, k; \mathbf{r}_t, \mathbf{v}_t, \mathbf{s}_t) = \exp\left(-\frac{1}{2}\mathbf{m}_t^T\mathbf{R}^T\mathbf{\Sigma}_t^{-1}\mathbf{R}\mathbf{m}_t - \frac{(k\Delta_z - z_t\|\mathbf{m}_t\|\tan\theta_t)^2}{2\sigma_{z,t}^2}\right) \tag{11}$$

with $\sigma_{z,t}$ ($\approx$ 235nm) modeling the axial blurring, z_t denoting the measured intensity, and $\mathbf{R} = \mathbf{R}(\phi_t)$ denoting a rotation matrix

$$\mathbf{R}(\phi_t) = \begin{pmatrix} \cos\phi_t & \sin\phi_t \\ -\sin\phi_t & \cos\phi_t \end{pmatrix}, \quad \mathbf{\Sigma_t} = \begin{pmatrix} \sigma_{\mathrm{m},t}^2(\theta_t) & 0 \\ 0 & \sigma_{\mathrm{min},t}^2 \end{pmatrix}, \quad \mathbf{m}_t = \begin{pmatrix} i\Delta_x - x_t \\ j\Delta_y - y_t \end{pmatrix},$$

$$\sigma_{\mathrm{m},t}(\theta_t) = \sigma_{\min,t} - (\sigma_{\min,t} - \sigma_{\max,t}) \cos\theta_t,$$

$$\tan\theta_t = \frac{\dot{z}_t}{\sqrt{\dot{x}_t{}^2 + \dot{y}_t{}^2}}, \quad \tan\phi_t = \frac{\dot{y}_t}{\dot{x}_t}, \quad -\pi < \phi_t, \theta_t \leq \pi.$$

In these formulae, each pixel (i, j, k) is assumed to correspond to a rectangular volume of dimensions $\Delta_x \times \Delta_y \times \Delta_z$ nm^3. The parameters $\sigma_{\max,t}$ and $\sigma_{\min,t}$ represent the amount of blurring and, at the same time, model object elongation along the direction of motion. For subresolution structures such as vesicles, $\sigma_{\min} = \sigma_{\max} \approx 80$nm, and for the elongated MTs, $\sigma_{\min} \approx 100$nm and $\sigma_{\max} \approx 300$nm. For the background level estimate b_t we use the average image intensity at time t, taking into account the contribution of object intensity values to the total image intensity (mainly formed by background structures with lower intensity) is negligible. For a typical 2D image of size $10^3 \times 10^3$ pixels containing a thousand objects, the number of object pixels is only about 1%.

3.3 Rao-Blackwellization and Likelihood

As mentioned in Section 2, in the case of high-dimensional state spaces (in our case n_x=10), the SIS becomes inefficient and leads to variance increase of the estimator. However, when the transition and observation models have an analytically tractable structure, the size of the state space can be drastically reduced by analytical marginalization of some of the state variables, also called Rao-Blackwellization [6]. In our case, for each realization of state variable $\mathbf{y}_t = (\mathbf{r}_t, \mathbf{v}_t, \mathbf{s}_t)^T$, we have a linear Gaussian transition and observation model for the intensity I_t. For such models the optimal solution can be obtained analytically using the Kalman filter. We therefore combine a PF to compute the distribution of the discrete states $\mathbf{y}_t$ with a bank of Kalman filters to compute exactly the distribution of the continuous state. With the factorization

$$p(\mathbf{y}_t, I_t | \mathbf{z}_{1:t}) = p(I_t | \mathbf{y}_t, \mathbf{z}_{1:t}) p(\mathbf{y}_t | \mathbf{z}_{1:t}), \tag{12}$$

the density $p(I_t|\mathbf{y}_t, \mathbf{z}_{1:t})$, which is Gaussian, can be computed analytically by applying the Kalman filter:

$$p(I_t | \mathbf{y}_t, \mathbf{z}_{1:t}) = \mathcal{N}(I_t | I_{t|t}, P_{t|t}), \tag{13}$$

with $\mathcal{N}(\cdot|\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2,

$$\begin{aligned} I_{t|t-1} &= (1-\alpha) I_{t-1|t-1}, & I_{t|t} &= I_{t|t-1} + K_t (Z_t - H_t I_{t|t-1}), \\ P_{t|t-1} &= (1-\alpha)^2 P_{t-1|t-1} + q_3 T, & P_{t|t} &= P_{t|t-1} - K_t H_t P_{t|t-1}, \\ S_t &= H_t P_{t|t-1} H_t^T + R_t, & K_t &= P_{t|t-1} H_t^T S_t^{-1}, \end{aligned}$$

and the vectors H_t and Z_t are formed as

$$H_t = (\dots, a_t(i, j, k; \mathbf{r}_t, \mathbf{v}_t, \mathbf{s}_t), \dots)^T, \quad Z_t = (\dots, z(i, j, k) - b_t, \dots)^T, \tag{14}$$

for all triplets $(i,j,k) \in C(\mathbf{x}_t)$, where $C(\mathbf{x}_t)$ is the region of pixels that are affected by the object with state $\mathbf{x}_t$ and is defined as $C(\mathbf{x}_t) = \{(i,j,k) \in \mathbb{Z}^3 : a_t(i,j,k;\mathbf{r}_t,\mathbf{v}_t,\mathbf{s}_t) > 0.1\}$. The covariance matrix of the measurement noise R_t models the Poisson noise, the main source of noise in fluorescence microscopy imaging, and is given by $R_t = \text{diag}[\ldots, h_t(i,j,k;\mathbf{x}_t), \ldots]$. The recursive Bayesian solution is applicable as long as the statistics of the measurement noise is known for each pixel. Thus, we need to estimate only $p(\mathbf{y}_t|\mathbf{z}_{1:t})$, using a PF, in a space of reduced dimension, which satisfies the alternative recursion

$$p(\mathbf{y}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{y}_{t-1}|\mathbf{z}_{1:t-1})p(\mathbf{z}_t|\mathbf{y}_t,\mathbf{z}_{1:t-1})p(\mathbf{y}_t|\mathbf{y}_{t-1}). \tag{15}$$

The likelihood $p(\mathbf{z}_t|\mathbf{y}_t,\mathbf{z}_{1:t-1})$ does not simplify to $p(\mathbf{z}_t|\mathbf{y}_t)$ because there is a dependency on past values through $I_{0:t}$. For conditionally linear models, we have $p(\mathbf{z}_t|\mathbf{y}_t,\mathbf{z}_{1:t-1}) = \mathcal{N}(b_t + H_t I_{t|t-1}, S_t)$ [6]. The variance of the importance weights for RB(M)PF is lower than for (M)PF [11]. Also, for the same performance, fewer MC particles are needed. This is because the dimension of $p(\mathbf{y}_t|\mathbf{z}_{1:t})$ is smaller than that of $p(\mathbf{x}_t|\mathbf{z}_{1:t})$. Another reason is that optimal algorithms are used in order to estimate the linear state variables.

3.4 Data-Dependent Sampling

Basic PFs [7,9], which use the proposal distribution $q(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{z}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$, usually perform poorly because too few samples are generated in regions where the desired posterior $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ is large. In order to construct a proposal distribution which alleviates this problem and takes into account the most recent measurements $\mathbf{z}_t$, we propose to transform the image sequence into probability distributions. True spots are characterized by relatively high intensities with convex profiles. Noise-induced local maxima typically exhibit a random distribution of intensity changes in all directions, leading to a low local curvature [3]. These two discriminative features (intensity and curvature) are used to construct an approximation of the likelihood $L(\mathbf{z}_t|\mathbf{x}_t)$, using the image data available at time t. For each object we use the transformation

$$\tilde{p}_m(\mathbf{r}_t|\mathbf{z}_t) = \frac{(G_\sigma * \tilde{z}_t(\mathbf{r}_t))^r \kappa_t^s(\mathbf{r}_t)}{\int_{C_{m,t}} (G_\sigma * \tilde{z}_t(\mathbf{r}_t))^r \kappa_t^s(\mathbf{r}_t) dxdydz}, \qquad \forall \mathbf{r}_t \in C_{m,t}, \tag{16}$$

where G_σ is the Gaussian kernel with scale σ, $\tilde{z}_t$ denotes a first-order interpolation of z_t, $C_{m,t}$ is the circular region (with radius defined by the covariance matrix of $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, e.g. 3-standard-deviation level) centered at the object position predicted from the previous time step, the curvature κ_t is given by the determinant of the Hessian matrix $\mathbf{H}$ of the intensity $\tilde{z}_t$:

$$\kappa_t(\mathbf{r}_t) = \det(\mathbf{H}(\mathbf{r}_t)), \qquad \mathbf{H}(\mathbf{r}_t) = \nabla \cdot \nabla^T \tilde{z}_t(\mathbf{r}_t), \tag{17}$$

and the exponents $r > 1$ and $s > 1$ weigh each of the features and determine the peakedness of the likelihood. Using this transformation, we define the new data dependent proposal distribution for object m as

$$\tilde{q}_m(\mathbf{y}_t|\mathbf{y}_{t-1},\mathbf{z}_t) = \tilde{p}_m(\mathbf{r}_t|\mathbf{z}_t)\mathcal{N}(\mathbf{v}_t|\mathbf{r}_t - \hat{\mathbf{r}}_{m,t-1}^{\text{MMSE}}, \mathbf{\Sigma}_\mathbf{v})\mathcal{N}(\mathbf{s}_t|\mathbf{s}_{m,t-1}^{\text{MMSE}}, q_2 T), \tag{18}$$

where the covariance matrix $\mathbf{\Sigma_v} = \mathrm{diag}[q_1 T, q_1 T, q_1 T]$. Contrary to the original proposal distribution, which fails if the likelihood is too peaked, (18) generates samples that are highly consistent with the most recent measurements in the predicted (using the information from the previous time step) regions $C_{m,t}$. A mixture of both proposal distributions gives excellent results:

$$q_m(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{z}_t) = \gamma p(\mathbf{y}_t|\mathbf{y}_{t-1}) + (1-\gamma)\tilde{q}_m(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{z}_t),$$

where $0 < \gamma < 1$. Compared to the regular one, this proposal distribution scales much better to smaller sample sizes.

3.5 Track Initialization and Management

The initialization of the proposed RBMPF can be done by manually specifying the objects of interest in the first frame, or by using a completely automatic initialization procedure [8]. The latter, also used for detection of newly appearing objects in subsequent frames, divides the image space into rectangular 3D cells and samples N_s particles according to importance function (16). The number of sampled particles in each cell represents the degree of belief in object birth. In cells (not containing any of the M existing objects) with counts larger than some threshold N_{td}, new tracks are initiated with initial mixture weights π_{bd}. The threshold N_{td} can be estimated experimentally and depends on N_s, cell volume and the number of bright spots in the image data.

Whenever objects pass close to one another, the object with the best likelihood score typically "hijacks" particles of nearby mixture components. This problem is partly solved using the MRF model for object interactions [8]. To better resolve the ambiguity in such situations, the Hough transform is used for each spatiotemporal ROI of 3-5 frames, $C_{m,t-\tau:t+\tau}$, to correctly model velocity changes. If object m passes close to a new object, the distribution $p_m(\mathbf{x}_t|\mathbf{z}_{1:t})$ becomes too diffuse in a few time steps and the reclustering procedure $(\{c'^{(i)}_t\}, M') = F(\{\mathbf{x}^{(i)}_t\}, \{c^{(i)}_t\}, M)$ [12] is performed to initiate new tracks. Merging of objects does not occur in our application and is therefore forbidden. If the mixture weight $\pi_{m,t}$ is below some predefined threshold level π_{td}, component m is removed from the mixture and the track terminated.

4 Experimental Results

4.1 Evaluation on Synthetic Data

The RBMPF was first evaluated and compared to the standard PF using synthetic but realistic 2D image sequences (20 frames of 512×512 pixels) of moving MT-like objects ($10-20$ objects per frame), generated according to (9) and (10), for different SNRs (Fig. 2) in a range around SNR=4, which has been identified by previous studies [4] as a critical level at which several popular tracking techniques break down. Object velocities ranged from 200 to 700 nm/sec, representative of published data [13]. The PFs used 300 samples per object and

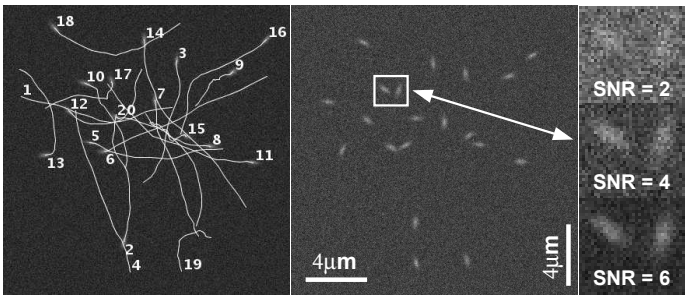

Fig. 2. Examples of synthetic images used in the experiments. The left image is a frame from one of the sequences (SNR=7) with the trajectories of the 20 moving objects superimposed, illustrating the motion patterns allowed by the linear state evolution model (9). The right image is a frame from another sequence (SNR=4), giving an impression of object appearance. The insets show zooms at different SNRs.

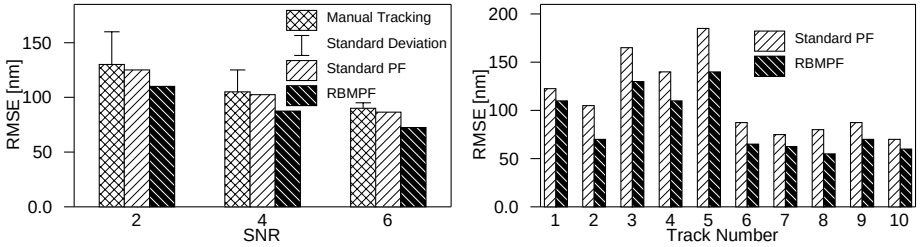

Fig. 3. Average RMSE (left) for different SNRs, and RMSE per track (right, SNR=4), for manual and automatic tracking in the described synthetic data sets.

the algorithm parameters were as follows: $\sigma_{\max} = 250$ nm, $\sigma_{\min} = 100$ nm, $\Delta_x = \Delta_y = 50$ nm, $r = 5$, $s = 1$, $q_1 = 3$, $\alpha = 10^{-3}$ and $q_2 = 0.04$, $q_3 = 1$, $T = 1$ sec. SNR is defined as the difference in intensity between the object I_o and background I_b, divided by the object noise, $\sigma_o = \sqrt{I_o}$ [4]. Tracks were initialized manually in desired regions of interest in the first frame, or automatically using the described procedure, giving similar results.

Having the ground truth for the synthetic data, we evaluated the accuracy of the tracking and compared the PF techniques and manual tracking (done by five independent expert observers). To quantify the localization error, we used the traditional root mean square error measure [3]. Figure 3 shows the RMSE in the object position estimates as a function of SNR for the manual tracking and using the PF techniques. The localization error of our algorithm is in the range of errors made by experts. The error bars represent interobserver variability and indicate that manual tracking performance degrades significantly for low SNRs. The figure also shows the reduction of the variance when the RBMPF is used, compared to the standard PF. The RBMPF was also evaluated on 3D synthetic image sequences, with 20 optical slices, $\sigma_z = 250$ nm, $\Delta_z = 200$ nm. The RMSE

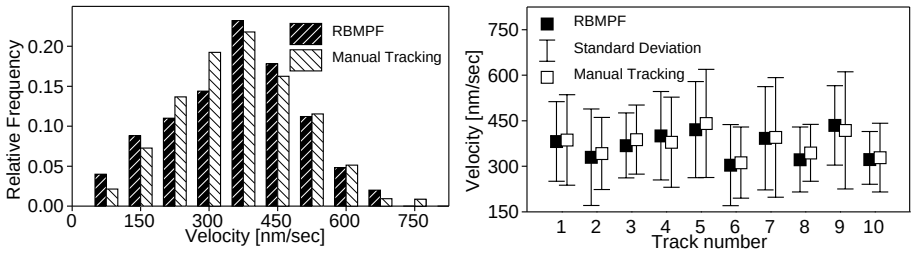

Fig. 4. Examples of velocity distributions (left) and velocity estimation for 10 representative MT objects (right) obtained with our RBMPF-based automatic tracking algorithm versus manual tracking applied to real fluorescence microscopy image sequences of growing MTs (data set in Fig. 1(a), SNR≈5).

in this case was higher (about 300-400 nm), due to the approximately three times lower optical resolution of the modeled imaging system in the axial direction, and the variance reduction by using RBMPF was not significant, due to the increase of the state space dimension for the PF from 6 to 9.

4.2 Evaluation on Real Data

The algorithm was also applied to real 2D fluorescence confocal microscopy image sequences acquired for MT dynamics studies [13]. Three representative data sets (of size 512×512 pixels, 30 time frames, examples of which are shown in Fig. 1) were preselected from larger volumes by manually choosing the regions of interest. In these experiments, the parameters of the algorithm were fixed to the same values as in the case of the synthetic data. Using the automatic track initiation procedure, the tracker simultaneously followed 10–30 spots during 3–20 consecutive frames until their disappearance.

Lacking ground truth for the real data, we evaluated the performance of our algorithm by comparison with manual tracking results from two expert cell biologists. Distributions of instant velocities estimated using our RBMPF-based algorithm versus manual tracking for SNR≈5 are presented in Fig. 4. Application of a paired Student t-test per track revealed no statistically significant difference between the results of our algorithm and that of manual tracking, for both expert human observers ($p \gg 0.05$ in all cases). The difference in average velocity (over 10 tracks) between automatic and manual tracking was less than 1%, for both observers. Our velocity estimates are also comparable to those reported previously based on manual tracking in the same type of image data [13]. Two different example visualizations of real data together with the results of tracking using our algorithm are given in Fig. 5.

5 Discussion and Conclusions

We have presented a Rao-Blackwellized marginal particle filter for tracking of multiple objects in molecular bioimaging data. The proposed approach contains

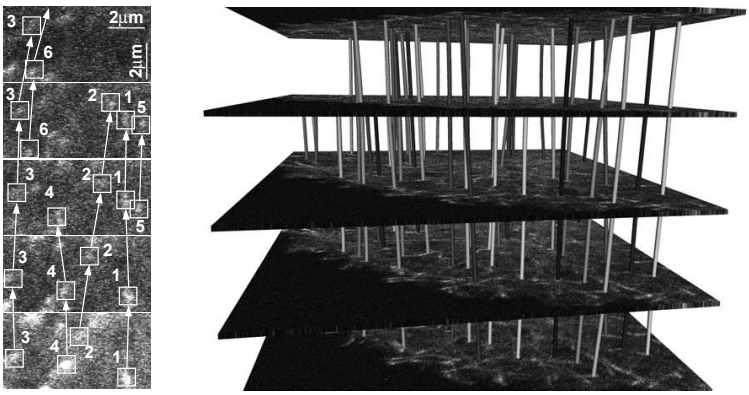

Fig. 5. Left: Results (six tracks) of automatically tracking MTs in the presence of photobleaching, illustrating the capability of our algorithm to capture newly appearing objects (tracks 5 and 6) and to detect object disappearance (track 4). It also shows the robustness of the algorithm in the case of closely passing objects (tracks 1 and 5). Right: Visualization of tracking results produced by our algorithm in the case of the real data of Fig. 1(a). Shown are five frames (time is increasing from bottom to top) with the trajectories rendered as small tubes.

several improvements over previous PF-techniques. Specifically, the robustness and reproducibility of the algorithm are improved by means of a new importance function for data-dependent sampling and by using marginalization for both the filtering distribution and selected variables from the state vector. These modifications reduce the number of MC samples required for tracking from $10^5 - 10^6$ [8] to $10^2 - 10^3$. Compared to existing deterministic approaches, which perform object detection prior to linking using non-Bayesian maximum likelihood or least squares estimators, the proposed estimator optimally exploits temporal information and prior information about the parameters, resulting in lower variance. As the experiments show, even in data with SNR as low as 2 (which is not uncommon in practice), our algorithm still yields reliable tracking results, whereas common frame-by-frame approaches break down at SNR$<$4–5 [4].

The results of the experiments on synthetic image data suggest that our algorithm is potentially more accurate than manual tracking by expert human observers. The experiments on real fluorescence microscopy image sequences from MT dynamics studies showed comparable performance. This is explained by the fact that in the latter experiments, we were limited to comparing distributions and averages (Fig. 4), which may conceal small local discrepancies, especially when the objects' velocities vary over time. Instant velocities were also analyzed per track but could not be quantitatively validated due to the lack of ground truth. Nevertheless, the results indicate that our algorithm may replace laborious manual procedures. Currently we are evaluating the method also for other biological applications to further demonstrate its advantages over current means of manual and automated quantification of subcellular dynamics. Our findings encourage use of the method to analyze complex biological image sequences not

only for obtaining statistical estimates of average velocity and life span, but also for detailed analyses of complete life histories.

References

1. Meijering, E., Smal, I., Danuser, G.: Tracking in molecular bioimaging. IEEE Signal Process. Mag. **23**(3) (May 2006) 46–53
2. Gerlich, D., Mattes, J., Eils, R.: Quantitative motion analysis and visualization of cellular structures. Methods **29**(1) (Jan. 2003) 3–13
3. Thomann, D., Rines, D.R., Sorger, P.K., Danuser, G.: Automatic fluorescent tag detection in 3D with super-resolution: Application to the analysis of chromosome movement. J. Microsc. **208**(1) (Oct. 2002) 49–64
4. Cheezum, M.K., Walker, W.F., Guilford, W.H.: Quantitative comparison of algorithms for tracking single fluorescent particles. Biophys. J. **81**(4) (Oct. 2001) 2378–2388
5. Sage, D., Neumann, F.R., Hediger, F., Gasser, S.M., Unser, M.: Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. IEEE Trans. Image Process. **14**(9) (Sep. 2005) 1372–1383
6. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer-Verlag, Berlin (2001)
7. Arulampalam, S.M., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2) (Feb. 2002) 174–188
8. Smal, I., Niessen, W., Meijering, E.: Advanced particle filtering for multiple object tracking in dynamic fluorescence microscopy images. In: Proceedings of the IEEE International Symposium on Biomedical Imaging. (Apr. 2007)
9. Isard, M., Blake, A.: CONDENSATION – Conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1) (Aug. 1998) 5–28
10. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing **10**(3) (Jul. 2000) 197–208
11. Klaas, M., de Freitas, N., Doucet, A.: Toward practical N^2 Monte Carlo: The marginal particle filter. In: Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI-05), AUAI Press (2005) 308–331
12. Vermaak, J., Doucet, A., Pérez, P.: Maintaining multi-modality through mixture tracking. In: Proceedings of the 9th IEEE International Conference on Computer Vision. (2003) 1110–1116
13. Stepanova, T., Slemmer, J., Hoogenraad, C.C., Lansbergen, G., Dortland, B., De Zeeuw, C.I., Grosveld, F., van Cappellen, G., Akhmanova, A., Galjart, N.: Visualization of microtubule growth in cultured neurons via the use of EB3-GFP (end-binding protein 3-green fluorescent protein). J. Neurosci. **23**(7) (Apr. 2003) 2655–2664
14. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. Part I: Dynamic models. IEEE Trans. Aerosp. Electron. Syst. **39**(4) (Oct. 2003) 1333–1364
15. Song, L., Hennink, E.J., Young, I.T., Tanke, H.J.: Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. Biophys. J. **68**(6) (Jun. 1995) 2588–2600
16. Khan, Z., Balch, T., Dellaert, F.: MCMC-Based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Machine Intell. **27**(11) (Nov. 2005) 1805–1819