Stochastic Geometry for Multiple Object Tracking in Fluorescence Microscopy

Paula Craciun, Josiane Zerubia

▶ To cite this version:

STOCHASTIC GEOMETRY FOR MULTIPLE OBJECT TRACKING IN FLUORESCENCE MICROSCOPY

Paula Crăciun   *Josiane Zerubia*

INRIA, Ayin team
2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France
Email: firstname.lastname@inria.fr

ABSTRACT

This paper proposes a framework for tracking multiple fluorescent objects in 2D + time video-microscopy. We present a novel batch-processing track-before-detect multiple object tracking approach based on a spatio-temporal marked point process model of ellipses. Our approach takes into account events such as births, deaths, splits and merges of objects which are motivated by the biological and physical considerations. We show the performance of the proposed model on synthetic biological data and a real total internal reflection fluorescence microscopy (TIRF) image sequence.

Index Terms— multiple object tracking, fluorescence microscopy, marked point process, stochastic geometry, RJM-CMC

1. INTRODUCTION

Microscope image and video quality has experienced a large increase in recent years and the analysis of microscope data is now commonplace in fields such as medicine or biological research. The high acquisition speed of the cameras allows the real time observation of dynamic cellular and sub-cellular processes. As such, this approach can generate vast amounts of data that have to be stored, processed and analyzed. Automatic and semi-automatic methods for information retrieval are of great importance to the human operator in this field.
Multiple object tracking is a tedious task in 2D + time video-microscopy data as the signal to noise ratio (SNR) is quite low [1] and the tracking difficulty is raised by the high number of objects, object appearances and disappearances, clutter and false alarms [2]. Sequential methods have been proposed to tackle this problem, including Kalman filters combined with Probabilistic Data Association [3, 4], Probability Hypothesis Density filters [5, 6] or Multiple Hypothesis Tracking [7, 8]. Although these methods can be used for real-time analysis, they can provide limited results due to their sequential nature,

i.e. the impossibility to correct past decisions based on current data.
In this paper, we propose an energy minimization-based approach for tracking multiple sub-cellular structures using a batch optimization scheme. Objects are modeled using ellipses. Objects are allowed to appear, disappear, split or merge. This design choice is motivated by the underlying biological and physical considerations of the highly complex protein-protein interactions. We design a novel marked point process model of ellipses and embed the widely known reversible jump MCMC into a simulated annealing scheme to simulate our model which leads to very good overall results.

2. BACKGROUND

The 3D image cube (2D + time) is modeled as a bounded set $K = [0, I_{h_{max}}] \times [0, I_{w_{max}}] \times \{1, \cdots, T\}$. A marked point process [9] of ellipses is considered on K, with the mark space M defined as $M = [a_m, a_M] \times [b_m, b_M] \times (-\frac{\pi}{2}, \frac{\pi}{2}] \times [0, L]$, where a_m, a_M and b_m, b_M are the minimum and maximum length of the semi-major and semi-minor axis respectively, $\omega \in]-\frac{\pi}{2}, \frac{\pi}{2}]$ is the orientation of the ellipse and $l \in [0, L]$ is its label. Thus, an ellipse u can be defined as $u = (c_h, c_w, t, a, b, \omega, l)$ and a marked point process of ellipses X is a point process on $W = K \times M$. While the semi-axes a and b and the orientation ω describe the physical properties of an ellipse, the label l is used as an identifier. Trajectories are extracted by grouping objects according to their label.
The distribution of a point process can be defined by its probability density function where the Poisson point process plays the analogue role of the Lebesgue measure on $\mathbb{R}^d$, where d is the dimension of the object space [10]. The Gibbs family of processes is used to define the probability density as follows:

$$f_\theta(X = \mathbf{X}|\mathbf{Y}) = \frac{1}{c(\theta|\mathbf{Y})} \exp^{-U_\theta(\mathbf{X},\mathbf{Y})} \qquad (1)$$

where:
- $\mathbf{X} = \{\mathbf{x_1} \cup \mathbf{x_2} \cup \cdots \cup \mathbf{x_t} \cup \cdots \cup \mathbf{x_T}\}$ is the configuration of ellipses, with $\mathbf{x_t}$ being the configuration of ellipses at time t;
- $\mathbf{Y}$ represents the 3D image cube;

- θ is the parameter vector;
- $c(\theta|\mathbf{Y})$ is the normalizing constant;
- $U_\theta(\mathbf{X}, \mathbf{Y})$ is the energy term.

Using the Maximun A Posteriori (MAP) criterion, the most likely configuration of objects corresponds to the global minimum of the energy:

$$X \in arg \max_{\mathbf{X} \in \Omega} f_\theta(X = \mathbf{X}|\mathbf{Y}) = arg \min_{\mathbf{X} \in \Omega}[U_\theta(\mathbf{X}, \mathbf{Y})]. \quad (2)$$

The energy function is divided in two parts: an external energy term, $U_{\theta_{ext}}^{ext}(\mathbf{X}, \mathbf{Y})$ which determines how good the configuration fits the input sequence, and an internal energy term, $U_{\theta_{int}}^{int}(\mathbf{X})$, which incorporates knowledge about the interaction between objects in a single frame and across the entire batch considered. The total energy can be written as the sum of these two terms:

$$U_\theta(\mathbf{X}, \mathbf{Y}) = U_{\theta_{ext}}^{ext}(\mathbf{X}, \mathbf{Y}) + U_{\theta_{int}}^{int}(\mathbf{X}). \quad (3)$$

The parameter vectors of the external and internal energy terms are θ_{ext} and θ_{int} respectively and $\theta = [\theta_{ext}, \theta_{int}]$. In our model, the parameter vector θ contains 4 parameters and is computed automatically (further details can be found in [11]).

3. ENERGY FORMULATION

3.1. Internal energy term

The internal energy term consists of a set of constraints meant for a correct detection of objects and to facilitate tracking. These constraints target the layout of the objects. For instance, geometric and physical consistency should be maintained (e.g. interpenetration among objects should be avoided).

The dynamic model. One property of tracking (as opposed to individual detections per frame) is that in most cases object trajectories are smooth. Hence, configurations in which objects exhibit a motion described by a dynamic model can be favored. This motion model, denoted by $d(\cdot, \cdot, \cdot)$, depends on the application. The dynamic model used throughout this work for microtubule scenarios is a constant velocity model [12]. Let $v \in \mathbf{x}_{t-1}$, $u \in \mathbf{x}_t$, $w \in \mathbf{x}_{t+1}$ be three objects, then the dynamic model can be written in terms of their positions $pos(v)$, $pos(u)$ and $pos(w)$ as follows:

$$d(v, u, w) := ||pos(v) - 2pos(u) + pos(w)||^2 \quad (4)$$

Note however that any motion model can be inserted into the model.

An energy term is designed to favor objects which follow a given motion model s.t. for an object u that exists at time t it is written as:

$$U_{dyn}^{int}(u) = \begin{cases} d(\cdot, u, \cdot) - dyn_0 & \text{if } \exists \ (v \in \mathbf{x}_{t-1} \text{ and } w \in \mathbf{x}_{t+1}) \\ & \text{s.t. } d(v, u, w) \leq dyn_0 \\ 0 & \text{otherwise} \end{cases}$$
$$(5)$$

where dyn_0 is a threshold that describes how much objects can deviate from the motion model but still be awarded. Note however that this motion model does not restrict the velocity of the objects in terms of magnitude. If the maximum velocity of objects is known a priori, it can be incorporated into the model by imposing a maximum distance between the objects v and u, respectively u and w.

The energy term that awards configurations which follow the dynamic model is the sum over all objects in the configuration:

$$U_{dyn}^{int}(\mathbf{X}) = \gamma_{dyn} \sum_{u \in \mathbf{X}} U_{dyn}^{int}(u). \quad (6)$$

This term favors the creation of objects where the data evidence is reduced but the dynamic model motivates the existence of an object.

Label persistence. In order to distinguish between distinct trajectories a label is added to the mark of each object. This label can be viewed as a trajectory identifier. Different labels mean different trajectories. Thus, the number of labels has to be kept closely related to the number of trajectories in the data set. Ideally, the large number of objects u scattered across the image sequence should be assigned to a rather small number of labels. In this regard, the set of labels present in a configuration $\mathbf{X}$ is constructed by $labels(\mathbf{X}) = \bigcup_{u \in \mathbf{X}} l(u)$, where $l(u)$ is the label of object u. Configurations where the number of distinct labels is small are favored in the following way:

$$U_{label}^{int}(\mathbf{X}) = -\gamma_{label}\left(\frac{1}{|labels(\mathbf{X})|}\right) \quad (7)$$

where $|labels(\mathbf{X})|$ represents the cardinality of the set.

Mutual exclusion. Handling object collision or overlapping in a given frame is a crucial aspect when detecting and tracking objects. In this model, an infinite penalty is attributed to any configuration that contains objects that overlap more than a given extent s. Thus, the probability of selecting such a configuration is zero. Denoting by $A(u_i, u_j) = \frac{Area(u_i \cap u_j)}{\min(Area(u_i), Area(u_j))}$ the area of intersection between the objects u_i and u_j, with $u_i, u_j \in \mathbf{x}_t$, the non-overlapping energy term can be defined as:

$$U_{overlap}^{int}(\mathbf{X}) = \gamma_o \sum_{t \in 0, \cdots, T} \sum_{1 \leq i \neq j \leq n(\mathbf{x}_t)} t_s(u_i, u_j) \quad (8)$$

with:

$$t_s(u_i, u_j) = \begin{cases} 0 & \text{if } A(u_i, u_j) < s \\ +\infty & \text{otherwise} \end{cases} \quad (9)$$

where $s \in [0, 1]$ corresponds to the amount of overlapping allowed by the model and $n(\mathbf{x}_t)$ is the number of objects in the configuration $\mathbf{x}_t$ at time t. Consequently, if two or more objects overlap more than the allowed amount s in a given frame t, only one of those objects will be detected.

The **internal energy term** is computed as the sum of its parts described in eq. 6, eq. 7 and eq. 8 weighted by the parameters γ_{dyn}, γ_{label} and γ_o respectively.

3.2. External energy

We use a statistical approach to determine whether an object is present or absent. In this sense, we compute the difference between two consecutive frames, D_t. This will result is a gray level image in which the intensity of a pixel represents the amount of change present at that location between the frames. The larger the amount of change (resp. the higher the intensity of a pixel), the more likely it is for a moving target to be present in that location. We use a sliding window u, which is translated over every pixel p. We have the following two hypotheses:

- **H$_0$**: the ellipse covers only the background without any target being present and denote the area with $\mathcal{A}$;

- **H$_1$**: the ellipse is placed in the center of a target. The area corresponding to the interior of a target is denoted $\mathcal{I}_u$, while the area corresponding to the exterior of a target is denoted $\mathcal{F}_u$.

We are interested in the following probabilities:

$$\begin{cases} P(H_0|D_t)\text{: The probability of } H_0, \text{ knowing the scene } D_t; \\ P(H_1|D_t)\text{: The probability of } H_1, \text{ knowing the scene } D_t. \end{cases} \tag{10}$$

We define the probabilities ratio in the following way:

$$R = \frac{P(H_1|D_t)}{P(H_0|D_t)} = \frac{P(D_t|H_1)}{P(D_t|H_0)} \frac{P(H_1)}{P(H_0)}, \tag{11}$$

where $P(H_0)$ is the probability that the window u covers only the background, independent of the observation and $P(H_1)$ is the probability of a target being located at the center of the window, independent of the observation.

We define the likelihood ratio:

$$\log \Lambda = \log \frac{P(D_t|H_1)}{P(D_t|H_0)} = \log(P(D_t|H_1)) - \log(P(D_t|H_0)). \tag{12}$$

Using the Neyman-Pearson lemma [13], the decision of whether a target is or not present at a given location is taken as follows:

$$\begin{cases} \log \Lambda > \tau, \text{ a target is present}; \\ \log \Lambda < \tau, \text{ a target is not present}. \end{cases} \tag{13}$$

Detection model. To compute $P(D_t|H_0)$ and $P(D_t|H_1)$ we will make the following simplifying assumptions [14, 15]:

- The intensity of each pixel in D_t is independent and uniformly distributed (A1);

- Each pixel within a target follows a Gaussian distribution of unknown parameters (mean and variance) (A2);

- Each pixel within the background follows a Gaussian distribution of unknown parameters (mean and variance) (A3);

- The variance within the target and the background is equal (A4).

On the one hand, it is essential to realize that these constraints are not very restrictive when a small sliding window is used, even in the presence of inhomogeneous background. On the other hand, it is clear that this model is not well adapted when the background presents a visible texture.

Using the assumptions (A1-3), we can write the logarithm of the probability of an area S as:

$$L(S) = -\frac{1}{2} \sum_{p_i \in S} \left[\log(2\pi\sigma_S^2) + \frac{(p_i - m_S)^2}{\sigma_S^2} \right]. \tag{14}$$

Hence, $P(D_t|H_0)$ and $P(D_t|H_1)$ can be calculated using eq. 14:

$$P(D_t|H_0) = P(\mathcal{A}) \Leftrightarrow \log(P(D_t|H_0)) = L(\mathcal{A})$$
$$P(D_t|H_1) = P(\mathcal{I}_u) \times P(\mathcal{F}_u) \Leftrightarrow$$
$$\Leftrightarrow \log(P(D_t|H_1)) = L(\mathcal{I}_u) + L(\mathcal{F}_u). \tag{15}$$

The logarithm of the likelihood ratio can be written as:

$$\log \Lambda = L(\mathcal{I}_u) + L(\mathcal{F}_u) - L(\mathcal{A}). \tag{16}$$

As the quantities m_S and σ_S in eq. 14 are unknown (here S stands for the areas $\mathcal{A}$, $\mathcal{I}_u$ and $\mathcal{F}_u$), a classical solution to solve this problem is to replace them by their maximum likelihood estimates $\hat{m}_S$ and $\hat{\sigma}_S$ respectively. After incorporating the assumption (A4) and making a few simplifications we obtain the following expression for eq. 16 [16]:

$$\log \Lambda = N_{\mathcal{I}_u} \hat{m}_{\mathcal{I}_u} + N_{\mathcal{F}_u} \hat{m}_{\mathcal{F}_u} - N_{\mathcal{A}} \hat{m}_{\mathcal{A}}, \tag{17}$$

where:
- $N_{\mathcal{I}_u}, N_{\mathcal{F}_u}, N_{\mathcal{A}}$ are the number of pixels in the areas $\mathcal{I}_u$, $\mathcal{F}_u$ and $\mathcal{A}$ respectively;
- $\hat{m}_{\mathcal{I}_u}, \hat{m}_{\mathcal{F}_u}, \hat{m}_{\mathcal{A}}$ are the estimated means of the areas $\mathcal{I}_u$, $\mathcal{F}_u$ and $\mathcal{A}$ respectively.

Choosing a threshold τ. We have obtained a simple and relatively fast pre-detection algorithm that can be described in the following way: For each time instance t, for each pixel of the difference image D_t, compute the generalized likelihood ratio given by eq. 17 for each sliding ellipse u. For each time t, we obtain a 2D matrix which we denote with $\mathbf{M}_{GLRT}^t$. An application-dependent threshold τ is applied to this matrix based on the desired probability of false alarms. A single threshold is manually chosen for an entire data set (details can be found in [16]).

External energy formulation. The external energy term for an object $u = (x_u, y_u, t_u, a_u, b_u, \omega_u, l_u)$ is written as follows:

$$U_{stat}^{ext}(u) = \begin{cases} -1 & \text{if } \mathbf{M}_{GLRT}^{t_u}(x_u, y_u) > \tau, \\ +1 & \text{if } \mathbf{M}_{GLRT}^{t_u}(x_u, y_u) < \tau. \end{cases} \tag{18}$$

The external energy term for a configuration $\mathbf{X}$ associated to the pre-detection result is:

$$U_{stat}^{ext}(\mathbf{X}) = \gamma_{stat} \sum_{u \in \mathbf{X}} U_{stat}^{ext}(u) \tag{19}$$

4. OPTIMIZATION

The energy is clearly not convex. The dependence caused by the high-order physical constraints is the main reason that drives the energy to be non-convex. The target distribution is the posterior distribution of $\mathbf{X}$, i.e. $\pi(\mathbf{X}) = f(\mathbf{X}|\mathbf{Y})$, defined on a union of subspaces of different dimensions. The most widely known optimization method for non-convex energy functions and an unknown number of objects is the reversible jump Markov Chain Monte Carlo (RJMCMC) sampler developed by [17]. RJMCMC uses a mixture of perturbation kernels to create tunnels through the walls of high energy. Standard perturbation kernels are used in our experiments. We embed a parallel implementation of the RJMCMC sampler [18, 19] in a simulated annealing scheme to find a minimizer of the energy function. A geometric law is used to decrease the temperature parameter of the simulated annealing.

5. RESULTS

The proposed approach is first tested on three synthetic biological image sequences. The image sequences have been generated with Icy [20], an online available toolbox for biological image sequences developed by the Quantitative Image Analysis Unit at the Pasteur Institute in Paris.

The sequences consist of 100 images, each 512×512 pixels, with approximately 35 objects per frame. The three sequences exhibit the signal to noise ratios (SNRs) of 20, 10 and 7 dB respectively. Difficult scenarios containing multiple intersecting trajectories are depicted in Table 1 and the quantitative tracking results are displayed in Table 2. The objects are considered to exhibit a directed motion.
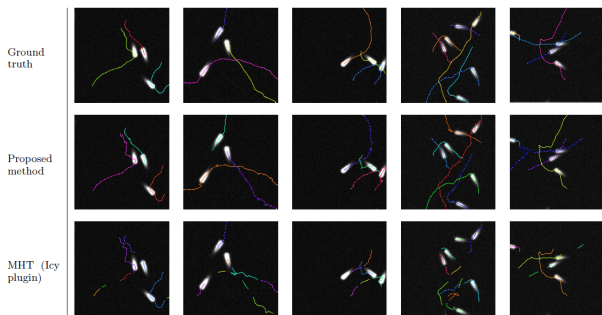

Table 1. Difficult scenarios with two or more crossing or by-passing trajectories in synthetic biological image sequences. First row: Ground truth trajectories. Second row: Trajectories obtained using proposed method closely resemble the ground truth trajectories. Third row: Trajectories obtained using MHT (Icy plugin) are highly fragmented.

The model has been applied on a real fluorescence microscopy sequence of 300 images of 512×256 pixels, by

Data set	Similarity between tracks			Similarity between detections		
	MHT	A. Milan et al. [2014]	Proposed algorithm	MHT	A. Milan et al. [2014]	Proposed algorithm
Seq. 1	43.7%	61.7%	**70.1%**	38.2%	57.8%	**72.4%**
Seq. 2	42.9%	60.9%	**70.0%**	35.9%	55.6%	**69.7%**
Seq. 3	43.6%	61.3%	**71.9%**	37.0%	56.9%	**61.5%**
Data set	Computation time / frame					
Seq. 1	1s (8CPUs)	2s (8CPUs)	4s (512CPUs)			
Seq. 2	1s (8CPUs)	2s (8CPUs)	6s (512CPUs)			
Seq. 3	2s (8CPUs)	3s (8CPUs)	9s (512CPUs)			

Table 2. Quantitative analysis of the detection and tracking results obtained using the built-in MHT tracker within Icy [20], the continuous energy minimization algorithm developed by [21] and the proposed method for the three synthetic biological image sequences. The proposed model outperforms current state of the art methods by more than 5%.
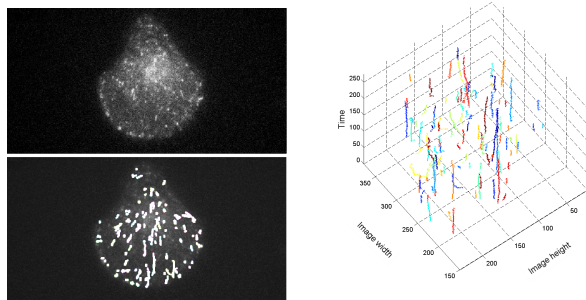

Fig. 1. Detection and tracking results (bottom left and right) on a real biological TIRF sequence of 300 images (top left; by courtesy of J. Salamero, PICT IBiSA, UMR 144 CNRS Institut Curie [22]). The visual assessment of the results reveals their very good quality.

courtesy of J. Salamero, PICT IBiSA, UMR 144 CNRS Institut Curie [22]. The TIRF image sequence shows the microtubule transport inside a cell. The tracking results are depicted in Figure 1. The visual assessment of the results is very good. Long, smooth tracks are identified which correspond to the motion patterns in microtubule transport.

6. CONCLUSIONS

In this paper we have proposed a novel spatio-temporal marked point process of ellipses to detect and track moving objects in fluorescent microscopy image sequences. This method can be applied to other types of imagery, e.g. remote sensing data. As the experimental results show, the novel method increases the performance by at least 5% compared to state of the art methods. The length of the image sequences does not influence the overall tracking accuracy. The proposed scheme could be extended to comprise various motion patterns which could be used for tracking objects with different behaviors.

7. REFERENCES

[1] S. van de Linde, M. Heilemann, and M. Sauer, "Live-cell super-resolution imagining with synthetic fluorophores," *Annual Review of Physical Chemistry*, vol. 63, no. 1, pp. 519–540, 2012.

[2] Y. Kalaidzidis, "Multiple object tracking in fluorescence microscopy," *Journal of Mathematical Biology*, vol. 58, no. 1-2, pp. 57–80, 2009.

[3] I. Smal, W. Niessen, and E. Meijering, "A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering," *Proc. ISBI*, pp. 264–267, 2008.

[4] W. J. Godinez and K. Rohr, "Tracking multiple particles in fluorescence time-lapse microscopy images via probabilistic data association," *IEEE Trans. Medical Imaging*, vol. 34, no. 2, pp. 415–432, 2015.

[5] J. Franco, J. Houssineau, D. Clark, and C. Rickman, "Simultaneous tracking of multiple particles and sensor position estimation in fluorescence microscopy images," *Proc. ICCAIS*, pp. 122–127, 2013.

[6] S.H. Rezatofighi, S. Gould, B.T. Vo, B.-N. Vo, K. Mele, and R. Hartley, "Multi-target tracking with time-varying clutter rate and detection profile: application to time-lapse cell microscopy sequences," *IEEE Trans. Medical Imaging*, vol. 34, no. 6, pp. 1336–1348, 2015.

[7] N. Chenouard, I. Bloch, and J.-C. Olivo-Marin, "Multiple hypothesis tracking in microscopy images," *Proc. ISBI*, pp. 1346–1349, 2009.

[8] L. Liang, H. Shen, P. De Camilli, and J.S. Duncan, "A novel multiple hypothesis based particle tracking method for clathrin mediated endocytosis analysis in fluorescence microscopy," *IEEE Trans. Image Proc.*, vol. 23, no. 4, pp. 1844–1857, 2014.

[9] M.N.M. van Lieshout and A.J. Baddeley, "Extrapolating and interpolating spatial patterns," in *Spatial Clustering Modelling*, A.B. Lawson and D.G.T. Denison, Eds. CRC Press/Chapman & Hall, 2002.

[10] X. Descombes, F. Chatelain, F. Lafarge, C. Lantuejoul, C. Mallet, M. Minlos, M. Schmitt, M. Sigelle, R. Stoica, and E. Zhizhina, *Stochastic Geometry for Image Analysis*, John Wiley and Sons, 2011.

[11] P. Craciun, *Stochastic geometry for automatic multiple object detection and tracking in remotely sensed high resolution image sequences*, Ph.D. thesis, Université de Nice - Sophia Antipolis, 2015.

[12] N. Chenouard, I. Smal, F. de Chaumont, et al., "Objective comparison of particle tracking methods," *Nature Methods*, vol. 11, no. 3, pp. 281–289, 2014.

[13] E.L. Lehmann and J. P. Romano, Eds., *Testing statistical hypotheses, 3rd edition*, Springer, 2008.

[14] V. Pagé, N. Proia, and G. Jubelin, "Maritime surveillance with the use of optical satellite images," *Proc. OCOSS*, 2010.

[15] N. Proia, *Surveillance maritime par analyse d'images satellitaires optiques panchromatiques*, Ph.D. thesis, Université des Antilles-Guyane, France, 2010.

[16] P. Craciun, *Stochastic geometry for automatic multiple object detection and tracking in remotely sensed high resolution image sequences*, Ph.D. thesis, INRIA-SAM / Université de Nice - Sophia Antipolis, France, available starting December 2015 at `https://team.inria.fr/ayin/publications-hal/`, 2015.

[17] P. Green, "Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[18] Y. Verdié and F. Lafarge, "Efficient Monte Carlo sampler for detecting parametric objects in large scenes," *Proc. ECCV*, vol. 7574, pp. 539–552, 2012.

[19] P. Crăciun and J. Zerubia, "Towards efficient simulation of marked point process models for boat extraction from high resolution optical remotely sensed images," *Proc. IGARSS*, pp. 2297 – 2300, 2014b.

[20] F. de Chaumont et al., "Icy: an open bioimage informatics platform for extended reproducible research," *Nature methods*, vol. 9, pp. 690–696, 2012.

[21] A. Milan, S. Roth, and K. Schindler, "Continuous energy minimization for multitarget tracking," *IEEE Trans. PAMI*, vol. 36, no. 1, pp. 58–72, 2014.

[22] A. Basset, J. Boulanger, P. Bouthemy, C. Kervrann, and J. Salamero, "SLT-LoG: A Vesicle Segmentation Method with Automatic Scale Selection and Local Thresholding Applied to TIRF Microscopy," *Proc. ISBI*, pp. 533–536, 2014.