Citation: Nguyen, T.T.D. and Shim, C. and Kim, W. 2021. Biological cell tracking and lineage inference via random finite sets. In: 18th IEEE International Symposium on Biomedical Imaging (ISBI), 13-16 Apr 2021, Nice, France. https://doi.org/10.1109/ISBI48211.2021.9433957

BIOLOGICAL CELL TRACKING AND LINEAGE INFERENCE VIA RANDOM FINITE SETS

† *Tran Thien Dat Nguyen* ⋆ *Changbeom Shim* ‡ *Wooil Kim*

School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Australia †⋆
Department of Computer Science and Engineering, Korea University, South Korea ‡
t.nguyen172@postgrad.curtin.edu.au † changbeom.shim@curtin.edu.au ⋆ wooilkim@korea.ac.kr ‡

ABSTRACT

Automatic cell tracking has long been a challenging problem due to the uncertainty of cell dynamic and observation process, where detection probability and clutter rate are unknown and time-varying. This is compounded when cell lineages are also to be inferred. In this paper, we propose a novel biological cell tracking method based on the Labeled Random Finite Set (RFS) approach to study cell migration patterns. Our method tracks cells with lineage by using a Generalised Label Multi-Bernoulli (GLMB) filter with objects spawning, and a robust Cardinalised Probability Hypothesis Density (CPHD) to address unknown and time-varying detection probability and clutter rate. The proposed method is capable of quantifying the certainty level of the tracking solutions. The capability of the algorithm on population dynamic inference is demonstrated on a migration sequence of breast cancer cells.

***Index Terms*—** Cell Tracking, Cell Lineage Inference, Track-By-Detection, Random Finite Set.

1. INTRODUCTION

Cell migration patterns are crucial in the study of cell reactions to certain conditions such as drug injections or changes in biochemical activities. Also, research on cell division gives insights into the developments of embryonic or tumors [1, 2]. Well-known approaches for tackling the cell tracking problem in the literature are track-by-detection [3–5] and model-based tracking [6–8]. On the other hand, from estimation perspective, the level of confidence governs the meaningfulness of tracking results. Hence it is necessary that a tracking framework can provide the uncertainty level together with the inferred results in order to fully characterise the tracking solution.

Cell tracking is an important practical applications of the multi-object tracking problem. The recent introduction of RFS theory [9] has opened up new pathways to address the problem from a rigorous mathematical perspective. Indeed, popular multi-object filters such as the CPHD, Multi-Bernoulli, GLMB, and Labeled Multi-Bernoulli (LMB) were derived from the RFS framework [9] and have been applied to track cells in [10–13]. In particular, the GLMB/LMB filters were developed to estimate objects trajectories via labeled RFS [14] and lineage of objects is also considered in [15]. Additionally, amongst the many approaches to multi-object tracking, the labeled RFS approach has demonstrated the capability for characterising confidence/uncertainty on the inferred results [16]. Furthermore, RFS-based filters can be formulated to handle tracking problems with unknown clutter rate and detection probability [17–19].

In this work, we propose a novel track-by-detection algorithm to study the migration pattern of cells where the clutter rate and detection probability of objects are unknown. In particular, this information is estimated via the robust CPHD filter [18] and then bootstrapped into the GLMB filter with objects spawning [15] to produce the posterior density of the multi-object state. This GLMB density provides statistics on the cell dynamic as well as the estimated trajectories with ancestral information via the GLMB trajectory estimator [20].

2. RFS-BASED CELL TRACKER

2.1. Multi-object Bayes Filter

We model an entire multi-object state as a random variable. Specifically, the set of objects is represented as a labeled RFS as $\mathbf{X} = \{(x_1, \ell_1), ..., (x_n, \ell_n)\}$ with $x_n \in \mathbb{X}$ and $\ell_n \in \mathbb{L}$, where $\mathbb{X}$ is the object state-space and $\mathbb{L}$ is a discrete label space. Similarly, $Z = \{z_1, ..., z_m\}$ denotes a measurement set with $z_m \in \mathbb{Z}$, where $\mathbb{Z}$ is the measurements space.

For a generic multi-object state X, $f(X_+|X)$ and $g(Z|X)$ represent the multi-object transition and measurement models, whereupon the posterior density $p(X_+|Z_+)$ of X can be computed via the Bayes filter as:

$$p(X_+) = \int p(X) f(X_+|X) \delta X, \tag{1}$$

$$p(X_+|Z_+) = \frac{p(X_+) g(Z_+|X_+)}{\int p(X_+) g(Z_+|X_+) \delta X_+}, \tag{2}$$

where '+' denotes the next time step.

* Corresponding author: C. Shim (changbeom.shim@curtin.edu.au)

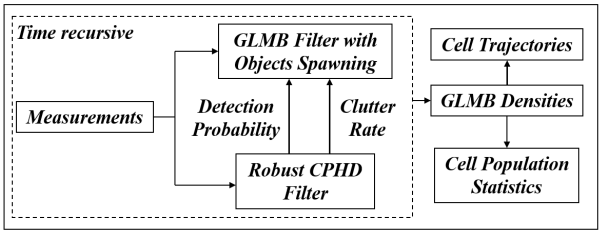


Fig. 1. Schematic of the proposed algorithm.

2.2. Estimating Clutter Rate and Detection Probability

Direct computation of the multi-object Bayes filter is prohibitively expensive, whereas the CPHD filter is the approximation of the Bayes filter via propagating the probability hypothesis density (PHD) and the cardinality distribution of the multi-object state. In this work, we adopt a robust CPHD filter for estimating the clutter rate and average detection probability. Then the estimated results are bootstrapped into a GLMB filter. The structure of the proposed algorithm is depicted in Fig. 1.

To accommodate unknown clutter rate, we follow the approach in [18], which considers objects of interest (cells) and clutter as two separate types of objects. Specifically, the filter uses $\mathbb{X}^{(1)}$ and $\mathbb{X}^{(0)}$ as the kinematic state-spaces of cells and clutter. A new hybrid state-space $\mathbb{X}^{(h)}$ of cell and clutter is defined as

$$\mathbb{X}^{(h)} = \left(\mathbb{X}^{(1)} \times \mathbb{X}^{(\Delta)}\right) \uplus \left(\mathbb{X}^{(0)} \times \mathbb{X}^{(\Delta)}\right), \qquad (3)$$

where $\mathbb{X}^{(\Delta)} = [0, 1]$ is the space of detection probability.

The multi-object state at the current time step evolves to a new set of objects, which consist of both *new objects* and *survival objects* at the next time step. Also, the birth rate and survival probability are related to how likely new births exist and objects survive. Especially, the probability density functions (PDFs) of new object state are proposed through the measurements from the previous time step, i.e., an adaptive birth model [21], while the PDFs of surviving objects are computed via single-object transition density. We assume cells follow the constant turn rate motion and the measurement model takes the form as in [18]. Moreover, the dynamic and measurement noises are modeled with Gaussian distributions, which are standard choices in practice. The detection probability of cells is modeled with Beta distribution and clutter is uniformly distributed over the entire image.

Given these models, the PHD and cardinality distribution on $\mathbb{X}^{(h)}$ then can be propagated recursively via Propositions 11 and 12 in [18]. Moreover, the cells and clutter are assumed to be statistically independent, from which the clutter rate $\bar{\kappa}$ and average detection probability of cells $\bar{p_D}$ can be extracted to be bootstrapped into GLMB filter being described subsequently.

2.3. Tracking Cells with the GLMB Filter

The GLMB filter in this step is based on the labeled RFS to estimate cell trajectories and lineage information. Given a set of cells at the current time step, the next (time step) new set of cells comprises of the independent births, surviving, and splitting cells. While the existing probability of independent births and survivals are governed by the birth rate and surviving probability, the probability of mitosis is governed by splitting probability. The PDFs of independent births are proposed via the measurements from the previous time step [21]. The PDFs of the surviving and splitting cells are computed from the current PDFs of the cells via the single-object transition densities. We also apply the standard multi-object measurement model in [15] and the noises are also assumed to be Gaussian.

The label of i^{th} birth cells appearing at time step k takes the form (k, i), and that of cells surviving at each time step is unchanged. Following the convention in [15], the number of new cell from division is only one and its label is denoted as $(\ell, k, 1)$, whereupon we can estimate the lineage in principle manner. Assuming that, the GLMB filter is to propagate the GLMB density [14]

$$p(\mathbf{X}) = \Delta(\mathbf{X}) \sum_{(I,\xi)\in\mathcal{F}(\mathbb{L})\times\Xi} \omega^{(I,\xi)} \delta_I(\mathcal{L}(\mathbf{X}))[p^{(\xi)}]^{\mathbf{X}}, \qquad (4)$$

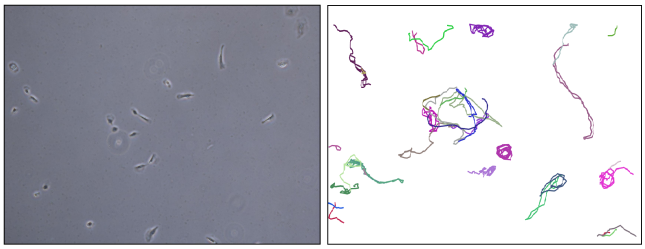
where $\Delta(\cdot)$ is the distinct label operator to ensure all labels of cells in $\mathbf{X}$ being unique, $\mathcal{F}(\cdot)$ is a collection of all finite subsets of the set in its argument, and $\delta_X(Y)$ is 1 if $X = Y$ and 0 otherwise. Moreover, $\xi \in \Xi$ is a history of association maps, i.e., $\xi = (\theta_1, ..., \theta_k)$, where θ_i is a function mapping track labels at time i to measurements indices at time i, and $\omega^{(I,\xi)}$ is the weight of hypothesis (I, ξ). Additionally, $p^{(\cdot)}(x, \ell)$ is the single-object PDF and $[p^{(\xi)}]^{\mathbf{X}} = \prod_{(x,\ell)\in\mathbf{X}} p^{(\xi)}(x, \ell)$.

The filter propagates the GLMB density in time via multi-object Bayes filter. The estimated clutter rate $\bar{\kappa}$ and the average detection probability $\bar{p}_D$ from the robust CPHD filter are used in the multi-object measurement model to compute the posterior density. The details on implementation is described in [15] which involve the uses of Gibbs sampler and a proposal density to select only significant hypotheses.

The time complexity of our method depends on those of CPHD and GLMB filters. Specifically, the CPHD filter with unknown clutter rate has $\mathcal{O}(M)$ time [18], and the GLMB filter mainly contributed by the Gibbs sampler entails $\mathcal{O}(TP^2M)$ time, where M is the number of received measurements, T is the number of requested hypotheses, and P is the number of available assignments. Readers are referred to [22] and [15] for more details of the GLMB filter complexity.

2.4. Statistical Inference on Cell Population

In this work, we use the trajectory estimator based on the association maps ξ of the estimated hypothesis at current time step k as proposed in [20] to provide estimation of the whole trajectories. In addition, given a GLMB density at current time step, the statistics of the cells population can be inferred via



(a) (b)

Fig. 2. (a) Snapshot of cells and (b) estimated cell trajectories.

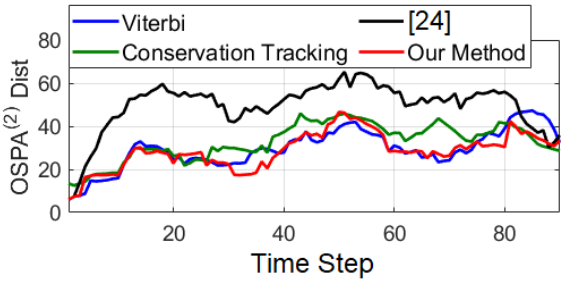


Fig. 3. OSPA$^{(2)}$ error for different methods.

the following equations [23]

$$\Pr(\text{total cells}{=}n) = \sum_{\xi\in\Xi}\sum_{I\subseteq\mathbb{L}} \delta_n\left[|I\cap\mathbb{L}_-|\right]\omega^{(I,\xi)}, \quad (5)$$

$$\Pr(\text{independent births }{=}n) = \sum_{\xi\in\Xi}\sum_{I\subseteq\mathbb{L}} \delta_n\left[|I\cap\mathbb{B}|\right]\omega^{(I,\xi)}, \quad (6)$$

$$\Pr(\text{mitotic events }{=}n) = \sum_{\xi\in\Xi}\sum_{I\subseteq\mathbb{L}} \delta_n\left[|I\cap\mathbb{S}|\right]\omega^{(I,\xi)}, \quad (7)$$

where $\mathbb{L}_-$ is the labels space at the previous time step, and $\mathbb{B}\subseteq\mathbb{L}$ and $\mathbb{S}\subseteq\mathbb{L}$ are the label spaces of independent births and splitting cells, respectively. On the other hand, the intensity of cell in the kinematic space is calculated as [23]

$$v(x,\ell) = \sum_{\xi\in\Xi} p^{(\xi)}(x,\ell)\sum_{I\subseteq\mathbb{L}} 1_I(\ell)\omega^{(I,\xi)}, \quad (8)$$

where $1_I(\ell)=1$ if $\ell\in I$ and 0 otherwise.

3. EXPERIMENTAL RESULTS

The proposed tracker is applied to a time-lapse image sequence of cancer cells with 90 frames (one frame per 15 minutes). The number of cells is time-varying due to births, cell division and deaths. A snapshot of the sequence is shown in Fig. 2(a). We focus on tracking the centroid of the cells and use the mitotic model in [24]. The parameters for the models and filters are set based on the characteristics of the dataset, the detection quality, and they are presented in Table 1.

In this experiment, we evaluate our method in comparison with the Viterbi linking [4], the conservation tracking [7] via ilastik software [25], and an RFS-based method in [24]. All approaches utilised the same detection set except for the conservation tracking; where, instead, we trained the classifiers for cell detection and division by using ilastik object detection and tracking pipelines. We adopted OSPA$^{(2)}$ metric (cut-off of 100 and window-length of 10) to measure errors between the estimated results and human annotations. This is a well-developed metric in the field of the multi-object tracking and can be interpreted as the time-averaged per-track error. We refer the reader to [26] for more details on the metric. Furthermore, we show statistical tracking solution, which can only be extracted by RFS-based methods. The estimated cell trajectories from our method are shown in Fig. 2(b).

Table 1. Parameters for the models and filters.

Parameters	Values
Surviving probability	0.999
Expected number of births per time step	4
Birth rate	0.03
Splitting probability	0.035
Standard deviation of process turn rate noise	$\pi/90$
Standard deviation of process velocity noise	5
Maximum number of GLMB components	10,000

In Fig. 3, the OSPA$^{(2)}$ errors indicate that the overall result of our method generally outweighs state-of-the-art algorithms. Specifically, the proposed method is superior to the method in [24] thanks to the trajectories estimator and the capability to estimate the unknown and time-varying clutter rate and detection probability. Although the conservation tracking algorithm performs relatively well, it cannot recover tracks after mis-detection hence worse performance than our approach. The Viterbi performance is competitive to our method given its capability to link the tracks via solving optimal paths problem and a state-space model. However, it cannot generate probabilistic inference on the tracking solution as in Bayesian approaches.

Fig. 4 depicts the estimation of cell population. Note that we omit the results of [24] due to the large performance gap comparing to those of other methods. Fig. 4(i) shows the estimated number of cells. Our algorithm provides the best cardinality estimation though all trackers inaccurately estimate the number of cells at around time steps 40 to 60 due to severe mis-detection. The differences between estimated number of division events and human annotations, the ground truth, are plotted in Fig. 4(ii). Here, positive/negative number indicates over/underestimate respectively. When cells clump (mis-detected) then split up, occasionally, our algorithm declares missed cells to be dead then it estimates additional division events to explain the extra measurements when cells are separated again. The conservation tracking tends to declare cells are dead when they are mis-detected then it declares false division events to explain new measurements. Meanwhile, the Viterbi algorithm tends to declare mis-detected cells to be dead hence low estimated cardinality but it provides better division detection. From this observation, it is shown that our algorithm has the balance of handling mis-detection and estimating

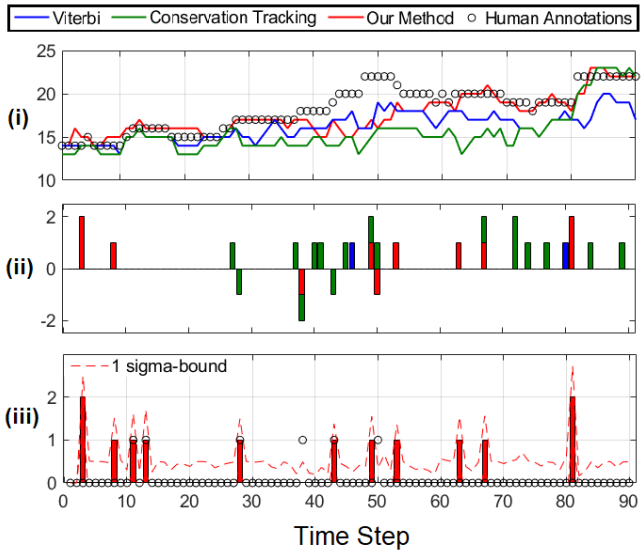


Fig. 4. The estimation of cell population.

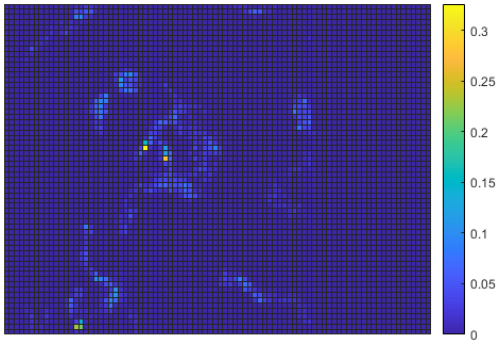


Fig. 5. Average cell intensity in location space.

division events to explain the measurements set. In details, Fig. 4(iii) shows the estimated number of division events at each time step provided by our method. The 1 sigma-bound curve is also shown to describe the uncertainty of the estimation, which can be quantified in Bayesian approach only.

We then demonstrate other capabilities of our approach in inferring the dynamic of the cell population. The average cell intensity in terms of location and velocity spaces are given in Figs. 5 and 6. These are calculated by summing cell intensities over all frames and then dividing by the number of frames. This probabilistic inference gives significant insights to cells migration pattern, in turn, their behaviours. For example, the intensity of the cells in velocity space shown in Fig. 6 indicated that there is no net drift in the x or y direction, i.e., no net migration of the cells. This observation cannot be made with the cell estimated trajectories. On the other hand, the properties such as stiffness of the environment [13] and overall motion direction of cells are reflected in Figs. 5 and 6. Also, Fig. 7 describes the cell lineage with that each color indicates a distinct family and the vertical connected lines indicate direct ancestral relationship. We assumed that cells are all independent new births at the first frame, i.e., there is no

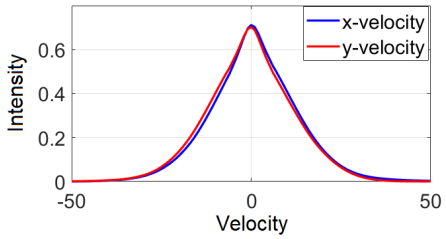


Fig. 6. Average cell intensity in velocity space.

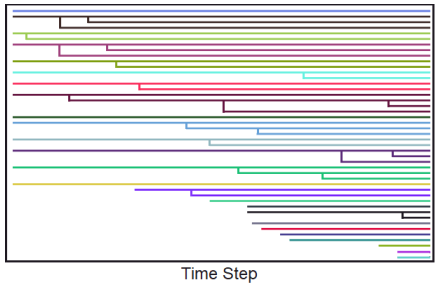


Fig. 7. Estimated cell lineage from our method.

initial ancestral relationship between cells. From this lineage map, the time and frequency of mitotic events can be inferred.

4. CONCLUSION

In this paper, we have proposed an RFS-based algorithm to estimate the cell trajectories with lineage in an environment where the clutter rate and detection probability of the cells are unknown and time-varying. The capability of the proposed approach in quantifying the certainty level of the tracking results has been also demonstrated. This study is one of a few first works on application of labeled RFS to cell tracking and the potential of this approach mainly remains unexplored. Future research directions are to extend the tracking capability using multi-sensor GLMB filtering to accommodate additional measurements sources, or by adapting the multi-scan GLMB filter to cell division model to enhance the estimation.

5. COMPLIANCE WITH ETHICAL STANDARDS

This is a study for which no ethical approval was required.

6. ACKNOWLEDGMENTS

This work was supported by Australian Research Council under Discovery Project DP160104662, Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A3A03039873), and under the framework of international cooperation program managed by the National Research Foundation of Korea (2020K2A9A1A01095894).

7. REFERENCES

[1] E. Meijering *et al.*, "Tracking in molecular bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 46–53, 2006.

[2] C. Zimmer *et al.*, "On the digital trail of mobile cells," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 54–62, 2006.

[3] O. Al-Kofahi *et al.*, "Automated cell lineage construction: A rapid method to analyze clonal development established with murine neural progenitor cells," *Cell Cycle*, vol. 5, no. 3, pp. 327–335, 2006.

[4] K. E. G. Magnusson *et al.*, "Global linking of cell tracks using the viterbi algorithm," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 911–929, April 2015.

[5] N. Chenouard *et al.*, "Multiple hypothesis tracking for cluttered biological image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2736–3750, 2013.

[6] M. Maska *et al.*, "Segmentation and shape tracking of whole fluorescent cells based on the chan–vese model," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 995–1006, June 2013.

[7] M. Schiegg *et al.*, "Conservation tracking," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 2928–2935.

[8] O. Debeir *et al.*, "Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes," *IEEE Transactions on Medical Imaging*, vol. 24, no. 6, pp. 697–711, 2005.

[9] R. P. S. Mahler, *Statistical multisource-multitarget information fusion*. Artech House, 2007.

[10] S. H. Rezatofighi *et al.*, "Multi-target tracking with time-varying clutter rate and detection profile: Application to time-lapse cell microscopy sequences," *IEEE Transactions on Medical Imaging*, vol. 34, no. 6, pp. 1336–1348, 2015.

[11] D. Y. Kim *et al.*, "A generalized labeled multi-bernoulli tracker for time lapse cell migration," in *2017 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2017, pp. 20–25.

[12] R. Hoseinnezhad *et al.*, "Visual tracking of numerous targets via multi-bernoulli filtering of image data," *Pattern Recognition*, vol. 45, no. 10, pp. 3625–3635, 2012.

[13] W. J. Hadden *et al.*, "Stem cell migration and mechanotransduction on linear stiffness gradient hydrogels," *Proceedings of the National Academy of Sciences*, vol. 114, no. 22, pp. 5647–5652, 2017.

[14] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, 2013.

[15] D. S. Bryant *et al.*, "A generalized labeled multi-bernoulli filter with object spawning," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6177–6189, Dec 2018.

[16] B.-N. Vo and B.-T. Vo, "A multi-scan labeled random finite set model for multi-object state estimation," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 4948–4963, Oct 2019.

[17] Y. G. Punchihewa *et al.*, "Multiple object tracking in unknown backgrounds with labeled random finite sets," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 3040–3055, June 2018.

[18] R. P. S. Mahler *et al.*, "CPHD filtering with unknown clutter rate and detection profile," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3497–3513, 2011.

[19] C.-T. Do *et al.*, "Tracking multiple marine ships via multiple sensors with unknown backgrounds," *Sensors*, vol. 19, no. 22, 2019.

[20] B.-N. Vo *et al.*, "Multi-sensor multi-object tracking with the generalized labeled multi-bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5952–5967, 2019.

[21] S. Reuter *et al.*, "The labeled multi-bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.

[22] B.-N. Vo *et al.*, "An efficient implementation of the generalized labeled multi-bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1975–1987, 2017.

[23] ——, "Labeled random finite sets and the bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, 2014.

[24] T. T. D. Nguyen and D. Y. Kim, "On-line tracking of cells and their lineage from time lapse video data," in *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, Oct 2018, pp. 291–296.

[25] S. Berg *et al.*, "ilastik: interactive machine learning for (bio)image analysis," *Nature Methods*, vol. 16, p. 1226–1232, Sep. 2019.

[26] M. Beard *et al.*, "A solution for large-scale multi-object tracking," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2754–2769, 2020.