PROTEIN TRACKING BY CNN-BASED CANDIDATE PRUNING AND TWO-STEP LINKING WITH BAYESIAN NETWORK

Mariia Dmitrieva[1], Helen L Zenner[2], Jennifer Richens[2], Daniel St Johnston[2] and Jens Rittscher[1] *

[1] Department of Engineering Science, University of Oxford, Oxford, UK
[2] Gurdon Institute, University of Cambridge, Cambridge, UK

ABSTRACT

Protein trafficking plays a vital role in understanding many biological processes and disease. Automated tracking of protein vesicles is challenging due to their erratic behaviour, changing appearance, and visual clutter. In this paper we present a novel tracking approach which utilizes a two-step linking process that exploits a probabilistic graphical model to predict tracklet linkage. The vesicles are initially detected with help of a candidate selection process, where the candidates are identified by a multi-scale spot enhancing filter. Subsequently, these candidates are pruned and selected by a light weight convolutional neural network. At the linking stage, the tracklets are formed based on the distance and the detection assignment which is implemented via combinatorial optimization algorithm. Each tracklet is described by a number of parameters used to evaluate the probability of tracklets connection by the inference over the Bayesian network. The tracking results are presented for confocal fluorescence microscopy data of protein trafficking in epithelial cells. The proposed method achieves a root mean square error (RMSE) of 1.39 for the vesicle localisation and α of 0.7 representing the degree of track matching with ground truth. The presented method is also evaluated against the state-of-the-art "Trackmate" framework.

Index Terms— tracking, probabilistic graphical models, bayesian network, biomedical imaging, convolutional neural network, tracklets, confocal microscopy, particle detection

1. INTRODUCTION

Biological particle tracking is essential in automated processing of a time lapse microscopy data and therefore being a subject of intense development. A vast number of tracking approaches have been developed and demonstrate high tracking accuracy [1], [2]. Nevertheless, tracking performance depends critically on the particular imaging conditions and the experimental set-up. Furthermore, most existing tracking approaches require tuning a large number of parameters or, in case of machine learning techniques, manually annotated dataset with defined tracks. To address these drawbacks we aim to reduce the number of adjustable parameters, and develop a framework that enhances the interpretability of the results. Our approach allows the user to adjust the computational model to specific experimental settings. Interpretability involves understanding of the track formation and how specific parameter settings affect the final result. End-to-end machine learning based approaches, which heavily depend on the training data and with a very large number of tunable hyperparameters, quite often do not satisfy this criteria.

Tracking approaches consist of two essential components: detection, and frame-to-frame object association (linking). Noisy data, signal loss due to photo-bleaching, presence of background fluorescence and other structures in the image [3] make the particle detection task in dynamic fluorescence microscopy challenging. Different approaches can be used for the detection [4], including the H-Dome transform-based detector [5], the spot-enhancing filter (SEF) [3] and the wavelet multiscale product [6]. Performance of these methods intensely declines when other structures appear on the image (e.g. membrane, Golgi). Lately, convolutional neural networks (CNN) are successfully exploited for the task (e.g. [7], [8], [9]). These approaches use a sliding window, UNet architecture or a recurrent network for the accurate particle detection.

The linking task can be approached by deterministic data association, where detection and linking implemented as two separate and independent processes (e.g. [10]). Probabilistic approaches, on the other hand, take uncertainty of the particle position and perform spatial-temporal filtering into account. Stochastic filtering (e.g. Kalman filter, Particle filters) requires prior knowledge of a suitable dynamic model and a spatial-temporal parameters of the target [11]. Therefore a priori assumptions on the dynamics of the particle movement need to be made. These approaches often are used in biological particle tracking in combination with other techniques (e.g. [12], [13], [14], [15]). Multiple Hypothesis Tracking and its modifications [16] build a logic scheme of hypotheses, updating them with each new measurement and therefore require high computational power. To achieve high accuracy

performance, a number of approaches use multi-stage linking methods (e.g. [17] [13]), where different techniques are combined to improve the linking task. Graphical models are exploited for data association [13, 18, 19], where vertices represent the detected particles and edges denote possible linking. An undirected grapchical model is used by Schiegg *et al.* [19] to incorporate prior beliefs from multiple classifiers, while in [18] authors use directed graphical model. Arasteh and colleagues [20] approach data association using a Hybrid Dynamic Bayesian Network. Recently, deep learning approaches are being used for data association. Recurrent neural network are employed for the linking task in [21], [9]. Denoising autoencoder (DAE) and score matching are used in [22] to solve the data association by learning dynamics of the data. These approaches allow to eliminate parameter tuning, however they require training data with known dynamics. "Trackmate" [23] provides a flexible solution to incorporate both, the particle detection and linking algorithms, for different type of data and commonly used for particle tracking.

This paper presents a particle tracking framework to support biological discovery of protein trafficking in epithelial cells and aims to provide flexible framework for the interpretation of estimated tracks. The protein trafficking data has been acquired using confocal fluorescence microscopy imaging which possess two main challenges for detection: 1) the cell membrane can be present in the same channel confounding protein vesicles detection; 2) multiple Golgi stacks can appear within a single cell. The aim of the detection approach is to provide a flexible solution, which would include both cases. And for the later, the Golgi stacks can appear as reasonably small spots which can be confused with vesicles. Thus, both membrane and Golgi stacks prohibit the efficient detection. To overcome these limitations, we propose to integrate vesicle candidate selection with a CNN based pruning. Particularly, a multi-scale spot enhancing filter (MSSEF) is used to detect a large number of candidate. Subsequently, a light weight CNN is employed for candidate selection. The proposed solution allows for performing candidate selection in real time and only requires very little training data.

Protein vesicle movement is unpredictable and depends on the particular experimental setting. Hence, the linking approach should be designed such that it is sensitive to the changes in vesicle movement without assuming any a priori knowledge of the dynamic model. The proposed linking approach is implemented by two-step linking process: first, short tracklets are being formed, then track linking is performed. Parameters of the tracklets are used to find optimal linking between them. We propose to use a probabilistic graphical model (PGM) for the track linking step. This approach allows us to model the probability of connecting a given set of tracklet into a larger tracks. With the help of PGMs we can create a flexible topology based on the provided parameter settings.

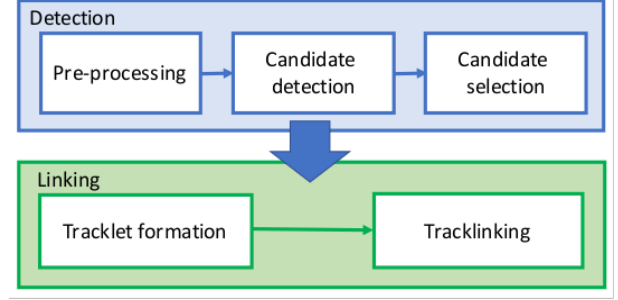We would like to highlight the following contributions: 1)

Fig. 1. Outline of the proposed tracking approach.

a robust candidate selection approach, where the employment of the light-weighted CNN allows reduction of tunable parameters providing state-of-the-art results; 2) application of a Bayesian Network to support track linking; 3) flexibility of the PGM topology facilitating an easy adaptation of the method to different experimental settings; and 4) comparative analysis with different topologies of PGMs. The paper is organised as follows: the tracking approach is described in Section 2, results and discussion are covered in Section 3, and conclusion with future work are outlined in Section 4.

2. METHODOLOGY

The proposed tracking method can be described by two major steps: vesicle detection and data association, as shown in Figure 1.

2.1. Vesicle detection

Vesicle detection is challenging due to the variance in their appearance, intensity and shape (see Figure 2 (e)). Moreover the membrane can be present at the same channel with the protein vesicles, and the Golgi stacks might look similar to the protein vesicles. The proposed vesicle detection is executed in three steps: (1) an pre-processing step is applied to each frame; (2) candidate detection which exploits MSSEF; (3) a candidate selection using a CNN classifier to eliminate spurious detections, Figure 2 (a) - (d).

During pre-processing, the background is subtracted from each frame. The background intensity is estimated from the consecutive m frames with a processing frame at the center, 1. The intensity of the following frame is normalised after the background subtraction.

$$I'_n = I_n - \min_{n-\frac{m}{2}...n+\frac{m}{2}} I \tag{1}$$

For the candidate detection, the MSSEF [13] is applied to a given image. This filter exploits a set of spot enhancing filters (SEF). The spot enhancing filter intensifies spots by convolving the original image $g(x,y)$ with a Laplacian-of-Gaussian (LoG) kernel $LoG(x,y,\sigma^k)$ and applying a basic threshold
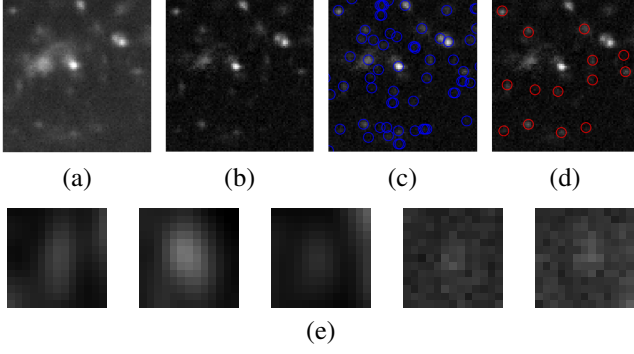
(a)   (b)   (c)   (d)

(e)

Fig. 2. Vesicle detection: (a) original frame; (b) frame after pre-processing step; (c) vesicle candidates; (d) detected vesicles (candidates selected by CNN); (e) difference in the vesicle appearance.

on the output image. In the multi-scale version, the SEF is applied sequentially for a set of LoG kernels, while the binary image from a previous step $b(x, y, \sigma^{k-1})$ masks the original image, i.e.

$$f(x, y, \sigma^k) = LoG(x, y, \sigma^k) * (b(x, y, \sigma^{k-1})g(x, y)), \quad (2)$$

where σ is a standard deviation of the LoG. The threshold for each iteration $T^{(k)}$ is defined by the mean intensity of the image $\mu_{intens}^{(k)}$, its standard deviation $\sigma_{intens}^{(k)}$ and a user defined constant c:

$$T^{(k)} = \mu_{intens}^{(k)} + c\sigma_{intens}^{(k)} \quad (3)$$

MSSEF allows to reduce the false detection related to the noise levels and to improve detection of the particles of different sizes. The list of vesicle candidates is formed from the local maxima coordinates of the MSSEF image. Parameter tuning for candidate selection is reduced to a set of standard deviations σ to define particle size and the threshold constant c for the particle brightness.

The simplicity of the vesicle shape provides opportunity to use a light weight CNN for the candidate selection with ROI of 16×16 pixels for each candidate. The proposed architecture is chosen to decrease the number of parameters without influencing the overall performance of the classifier. The network consists of only two convolutional layers followed by a max-pooling and a pair of fully connected layers at the final stage, Figure 3. This light architecture allows to reduce amount of training data to a number of hundred samples and provides fast candidate selection process.

2.2. Linking algorithm

The data association of the detected vesicles is implemented in two steps. Firstly, a set of short tracklets are formed, then these are connected using a probabilistic graphical model.
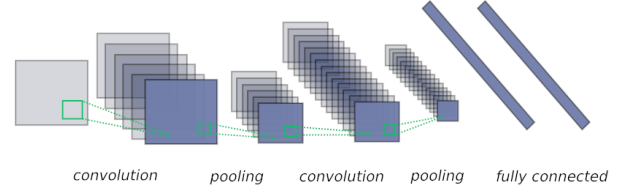

convolution   pooling   convolution   pooling   fully connected

Fig. 3. Architecture of the light weight CNN for candidate selection
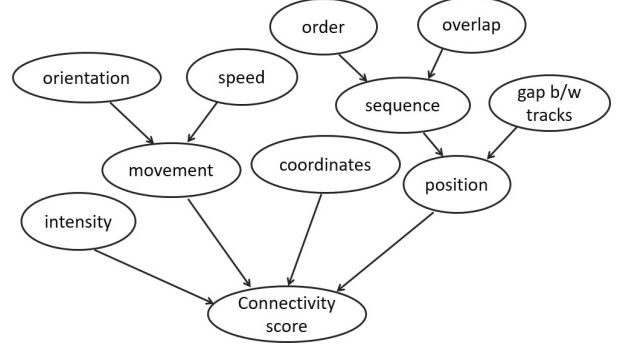

Fig. 4. Topology of the proposed BN for connectivity score evaluation.

The tracklet formation takes distance between the detected vesicles into account. The association of vesicles into tracklets is obtained by Hungarian algorithm, which provides a base for the data association. To create a reliable tracklets, the number of the skipped frames should be very small. The acceptable distance for the connection is also limited and the assignments with a large distances are removed. The final constraint on the tracklets is their length. Once the length of a tracklet reaches the assigned limit, the tracklet is removed from the active tracking. Short tracklets with reliable linking distance and small time gaps provide a stable and reliable fragments for track linking. This step requires three parameters: acceptable number of skipped frames (temporal gap), distance limit and maximum tracklet length.

Track linking of the assembled tracklets is based on their connectivity score. The score represents a probability of the tracklet pair to be connected. It is defined by an inference over a Bayesian network (BN). This probabilistic graphical model with a directed acyclic graph structure displays a set of discrete variables with their conditional dependencies. The variables represent similarity of the tracklet pair for different basis (parameters of the tracklets). The topology of the BN is built based on the known protein behavior and takes into account spacial and temporal location of the tracklets in relation to each other, their similarity in motion dynamics and particle appearance, Figure 4.

The connectivity score is a binary variable, where $True$ relates to tracklets being connected. It has four parents: inten-
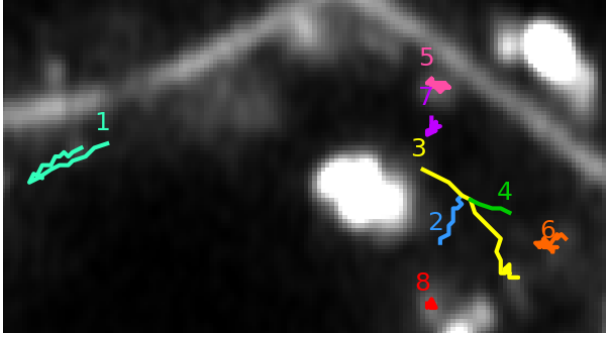
Fig. 5. Example of the tracks obtained by the proposed two-step data association approach.

(a)   (b)

Fig. 6. Example of the original protein trafficking dataset: (a) protein cargo channel of the spinning disc microscope, (b) airyscan with a single channel containing both membrane and protein cargo

sity, movement, coordinates and position. The intensity node represents the similarity in average intensity for the last frame of the tracklet #1 and the first frame of the tracklet #2. As it was mentioned earlier, intensity of the moving vesicles can vary and therefore this parameter will not be given high priority in conditional probability table. The coordinate node evaluates similarity in location between the last frame of the tracklet #1 and the first frame of the tracklet #2. The Euclidean distance between the points is used as a reference for the node outcome. This parameter is one of the most important and has a large impact on the connectivity score. The movement node defines similarity in the motion dynamic of the tracklet pairs and has two parents - orientation of the movement and its speed. The position node describes temporal location of the tracklets with two parents, which define sequential component and the temporal gap between them. The sequence node has two parents - order of the tracklets and their possible overlap in temproal domain. This node is essential for connecting tracklets in the right order. Gap between tracks represents the number of frames between the tracks and allows to priorities connections with a smaller time gap. The inference over the BN finds probability of the tracklet pair to be connected, P_c:

$$P_c = P(CS = true|I, Or, Sp, Od, Ov, G), \qquad (4)$$

where I is difference in intensity, Or - in orientation, Sp - speed difference, Od - order of the tracklets, Ov - presence of the overlap and G is a temporal gap between the tracklets.

The conditional probability table for each node can either be assigned manually based on the acquired knowledge about the vesicle movement, or trained with available data. The tracklet pairs with connectivity score higher than a given threshold will be accepted for consideration. For the conflict cases, where there are more than one possible connections, the pair with highest probability is selected. Figure 5 illustrates final tracks of a complex case and shows advantage of the proposed technique. Tracks #2 and #4 appear at the similar position and timing of the same location in track #3. The only difference between the tracklets is their orientation. The
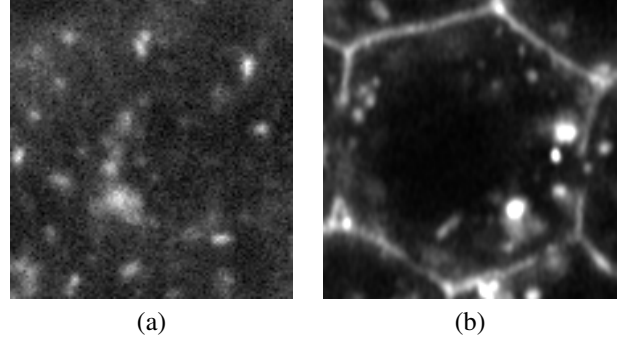
topology of the network allows to prioritise the orientation and build an appropriate association. On the other hand, track #1 illustrates large changes in the vesicles orientation, which where captured by the same BN and all the tracklets were connected to a single track as there were no other candidates with matching orientation.

The proposed Bayesian Network in implemented using pgmpy package [24]. The topology of the network can be adjusted based on the data dynamics and vesicle appearance. There are parameters which can be tuned at the track linking step. These include limits for temporal gap between the tracklets, coordinates deference limit, speed and orientation variation, intensity changes and a final threshold for tracklets connection.

The proposed linking solution is not designed to be executed in real time as it includes two-step data association. The run time heavily depends on the number of tracklets at the track linking step. To reduce the computation, the number of tracklets for comparison is limited by their spatial-temporal location, i.e. only connection between the tracklets which are close to each other is evaluated.

3. RESULTS AND DISCUSSION

Performance of the proposed tracking method is evaluated on the original data from the confocal fluorescence microscopy imaging of protein trafficking in epithelial cells of the drosophilae egg chamber. The tracking approach is illustrated for two different set-ups: spinning disk confocal microscope with two channels presenting membrane in one channel and protein cargo in another; and airyscan confocal microscope with a single channel which contains both, membrane and protein cargo, Figure 6.

Training data for the CNN classifier is collected from seven complete frames (each frame contains multiple cells) and both microscopes(four from airyscan and three from

Table 1. Performance of the proposed tracker and "Track-mate" framework.

Topology	RMSE	α	β	JSC
proposed method	**1.39**	**0.70**	**0.49**	**0.46**
"Trackmate"	7.39	0.14	0.08	0.2

Table 2. Performance of the proposed tracker for the spinning disk microscopy imaging with variation in BN topology.

Topology	RMSE	α	β	JSC
Full version	**1.63**	**0.69**	**0.56**	**0.57**
No speed	1.66	0.63	0.50	0.54
No orientation	**1.63**	**0.69**	**0.56**	0.55
No intensity	**1.63**	**0.69**	**0.56**	**0.57**
No time-gap	2.71	0.63	0.43	0.47

Table 3. Performance of the proposed tracker for the airyscan microscopy imaging with variation in BN topology.

Topology	RMSE	α	β	JSC
complete version	**1.16**	0.68	0.39	0.33
without speed	**1.16**	0.69	0.39	0.28
without orientation	**1.16**	**0.72**	**0.43**	**0.36**
without intensity	**1.16**	0.68	0.39	0.33
without time-gap	1.83	0.66	0.30	0.30

spinning disk microscopes). The probabilities for the BN are defined manually and therefore training is not required. The tracker is tested on six image sequences, where three sequences are obtained by the spinning disk microscope and three by the airyscan microscope. Each sequence contains 50 frames and images protein trafficking in a single cell. The tracker performance is compared with the performance of "Trackmate" [23] on the same data, where the detection is based on LoG detection and linking is performed by the [17], Table 1. The performance is quantified by the scores proposed at the particle tracking challenge [2]. Localisation of the vesicles is evaluated by the root mean square error (RMSE), which represents the overall localisation accuracy in the paired tracks. Performance of the linking method is quantified by the Jaccard similarity coefficient JSC and by the degree of matching between ground truth (GT) and estimated tracks with α (without taking into account spurious tracks) and β (penalising for the spurious tracks) [2]. The LoG based particle detection is not sufficient for the challenging data and causes low scores for the "Trackmate" framework. The results illustrate advantage of the proposed method.

The performance of the proposed tracker is also quantified for different BN topologies. The coordinate similarity and sequential order of the pairing tracks are considered as essential

components and were not changed. Tables 2 and 3 illustrate quantitative results of different BN topologies for spinning disk and airyscan microscopy imaging respectively. It can be concluded that the BN topology influences quality of the final result and therefore its flexibility is beneficial for adaptation of the approach for a different experimental settings.

Variation in the particle appearance and their similarity with Golgi stacks results in a large number of spurious tracks. It is well illustrated in Table 3 by the difference between α and β scores, where the large number of spurious tracks is caused by the cluttered and noisy background. The quantitative evaluation of the approach illustrates promising results and shows benefits from the employment BN for track linking.

Interpretability of the detection result is achieved by the candidate selection, where the candidates are obtained for a range of particle sizes and intensities and the selection process is based on the expected appearance of the particles. The data association interpritability is based on the BN topology and its relation to the particle movement. In total, there are two main tunable parameters for the candidate detection and a trained CNN network for the selection process. The first step of the candidate selection involves three parameters, while the BN topology and its variables can include up to six parameters. The conditional probability tables can be defined manually or obtained via BN training. The BN topology can be chosen from the available variations. It allows adaptation of the solution for different data, but keeps the number of the tunable parameters at its minimum.

4. CONCLUSION

This work presents a novel tracking approach to support exploratory biological studies of protein trafficking in epithelial cells. Targeting flexible framework, parameters reduction and interpretability of the results, we have employed candidate selection process with MSSEF and CNN-based classifier for the protein vesicle detection and two-step data association approach, where track linking is based on the inference over the Bayesian Network. This approach allows vesicle detection in a presence of other structures and provide robust data association.

In future work, the proposed linking approach can benefit from the CNN feature based similarity descriptor. It would allow improvement in the description of the vesicle appearance and provide opportunity to extend the work to the cell tracking. Furthermore, at the step of tracklet formation, the combinatorial optimization algorithm can be replaced with a different solution to incorporate more complex tracklet formation. It would eliminate a potential error in data association which can appear for a large tracklet length or a dense populated environment. Finally, the modification of the BN topology would benefit from an interactive interface, which provides an easy network modification and will be implemented as part of the vesicle tracking plug-in.

5. REFERENCES

[1] I Smal and E Meijering, "Quantitative comparison of multiframe data association techniques for particle tracking in time-lapse fluorescence microscopy," *Medical Image Analysis*, vol. 24, no. 1, pp. 163 – 189, 2015.

[2] N Chenouard, I Smal, F de Chaumont, M Maska, and et al, "Objective comparison of particle tracking methods," *Nature Methods*, vol. 11, 2015.

[3] D Sage, F R Neumann, F Hediger, S M Gasser, and M Unser, "Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics," *IEEE Tran. on Image Processing*, vol. 14, no. 9, pp. 1372–1383, 2005.

[4] I Smal, M Loog, W Niessen, and E Meijering, "Quantitative comparison of spot detection methods in fluorescence microscopy," *IEEE Trans. on Medical Imaging*, vol. 29, no. 2, 2010.

[5] L Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. on Image Processing*, vol. 2, no. 2.

[6] J-C Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognition*, vol. 35, no. 9, pp. 1989 – 1996, 2002.

[7] T Wollmann, C Ritter, J N Dohrke1, J-Y Lee, R Bartenschlager, and K Rohr, "Detnet: deep neural network for particle detection in fluorescence microscopy images," in *ISBI*, 2019, pp. 517–520.

[8] J M Newby, A M Schaefer, P T Lee, M G Forest, and S K Lai, "Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D," *Proceedings of the National Academy of Sciences*, vol. 115, no. 36, pp. 9026–9031, 2018.

[9] P R Gudla, K Nakayama, G Pegoraro, and T Misteli, "Spotlearn: Convolutional neural network for detection of fluorescence in situ hybridization (FISH) signals in high-throughput imaging approaches," in *CSH Symposia on Quantitative Biology*, 2017.

[10] K Shafique and M Shah, "A noniterative greedy algorithm for multiframe point correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, 2005.

[11] Y Salih and A S Malik, "Comparison of stochastic filtering methods for 3D tracking," *Pattern Recognition*, vol. 44, no. 10, 2011.

[12] I Smal, E Meijering, K Draegestein, N Galjart, and et al, "Multiple object tracking in molecular bioimaging by rao-blackwellized marginal particle filtering," *Medical Image Analysis*, vol. 12, no. 6, pp. 764 – 777, 2008.

[13] A Jaiswal, W J Godinez, R Eils, M J Lehmann, and K Rohr, "Tracking virus particles in fluorescence microscopy images using multi-scale detection and multiframe association," *IEEE Trans. on Image Processing*, vol. 24, no. 11, pp. 4122–4136, 2015.

[14] W J Godinez, M Lampe, S Wörz, B Mller, R Eils, and K Rohr, "Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences," *Medical Image Analysis*, vol. 13, no. 2, pp. 325 – 342, 2009.

[15] M Liu, Y He, Y Wei, and P Xiang, "Plant cell tracking using kalman filter based local graph matching," *Image and Vision Computing*, vol. 60, pp. 154 – 161, 2017.

[16] C Kim, F Li, A Ciptadi, and J Rehg, "Multiple hypothesis tracking revisited," in *ICCV*, 2015, pp. 4696–4704.

[17] K Jaqaman, D Loerke, M Mettlen, H Kuwata, S Grinstein, S L Schmid, and G Danuser, "Robust single-particle tracking in live-cell time-lapse sequences," *Nature Methods*, vol. 5, pp. 695 – 702, 2008.

[18] S Akram, J Kannala, L Eklund, and J Heikkilä, "Cell tracking via proposal generation and selection," *arXiv:1705.03386*, 2017.

[19] M Schiegg, P Hanslovsky, C Haubold, U Koethe, L Hufnagel, and F A Hamprecht, "Graphical model for joint segmentation and tracking of multiple dividing cells," *Bioinformatics*, vol. 31, no. 6, pp. 948–956, 2014.

[20] A Arasteh, B Vahdat, and S Reza, "Multi-target tracking of human spermatozoa in phase-contrast microscopy image sequences using a hybrid dynamic bayesian network," *Scientific Reports*, vol. 8, pp. 2045–2322, 2018.

[21] R Spilger, T Wollmann, YQiang, A Imle, J Y Lee, B Müller, O T Fackler, R Bartenschlager, and K Rohr, "Deep particle tracker: Automatic tracking of particles in fluorescence microscopy images using deep learning," in *DLMIA and ML-CDS*, 2018.

[22] I Smal, Y Yao, N Galjart, and E Meijering, "Facilitating data association in particle tracking using autoencoding and score matching," in *ISBI*, 2019, pp. 1523–1526.

[23] J-Y Tinevez, N Perry, J Schindelin, G M Hoopes, G D Reynolds, E Laplantine, S Y Bednarek, S L Shorte, and K W Eliceiri, "Trackmate: An open and extensible platform for single-particle tracking," *Methods*, vol. 115, pp. 80 – 90, 2017, Image Processing for Biologists.

[24] A Ankan and A Panda, "Pgmpy: Probabilistic graphical models using python," in *the 14th python in science conference*, 2015, pp. 6–11.