Multiple-target tracking of 3D fluorescent objects based on simulated annealing

Victor Racine, Ariane Hertzog, Jacqueline Jouanneau, Jean Salamero, Charles Kervrann, Jean-Baptiste Sibarita

▶ To cite this version:

HAL Id: hal-02757151
https://hal.inrae.fr/hal-02757151

Submitted on 3 Jun 2020

MULTIPLE-TARGET TRACKING OF 3D FLUORESCENT OBJECTS BASED ON SIMULATED ANNEALING

Victor Racine[1], Ariane Hertzog[2], Jacqueline Jouanneau[2], Jean Salamero[1], Charles Kervrann[3]
and Jean-Baptiste Sibarita[1]

[1] Imaging Center / UMR144, [2] Morphogenèse cellulaire et progression tumorale,
Institut Curie 26 rue d'Ulm, 75005 Paris, France.
[3] IRISA-INRIA Rennes / INRA - MIA, Campus de Beaulieu, 35042 Rennes Cedex, France.
JBS@curie.fr

ABSTRACT

This paper presents a framework for tracking fluorescent objects imaged with 3D+time video-microscopy. The proposed technique solves the NP-hard problem of the multi-frame object correspondence. For this purpose, we use the simulated annealing algorithm for its flexibility for feature correspondence and its ability to give results in a very short time. Our approach takes into account events like "split", "merge", "birth" and "death". These "evolutionary events" are motivated by biological and physical considerations. We demonstrate the performance of the proposed algorithm on images of carcinoma NBT-II cells expressing fluorescent LAR-PTP, a tyrosine phosphatase possibly involved in tumor progression.

1. INTRODUCTION

Rapid 3D + time (4D) microscopy has enormous potential for increasing our understanding of fast biological processes. However, to analyze the dynamics of the collected data, visualization and quantification using appropriate tools are required and still remains a bottleneck.

Multi-target tracking is a difficult task in computer vision and a large number of methods have been proposed those recent years. In most cases, this implies a two step process: first, objects are independently detected in each frame, and secondly, the detected objects are linked frame after frame to define object tracks. Therefore a prior motion model must be specified to impose some constraints and a specific strategy is needed to find the "best" solution in some sense, among the huge amount of possible solutions for matching. In the statistical framework, the two best known approaches are the joint probabilistic data-association filter (JPDAF) [1] and the multiple hypothesis tracker (MHT) [2]. In biology, such framework has been successfully proposed [3], but their major drawbacks are the large number of parameters and the assumption about probability distribution which does not necessarily hold [4].

Alternatively, we choose to minimize an objective criterion aiming to estimate a graph which describes the optimal set of object trajectories. Unlike stochastic filtering-based methods, we consider the image sequence as a volume in which we want to solve the "multiframe" correspondence problem. For this purpose, the detected objects at each frame can be connected by a grouping principle: features are linked using "minimal path" techniques applied to spatio-temporal data as described in [5]. The motion correspondence problem can be cast into a global energy minimization problem. We use the assumption that tracks are supposed to be smooth, a constraint based on the distance between successive detected points [6, 7]. This assumption can be encoded in a probability distribution in a bayesian framework. The optimization problem can be solved using greedy approaches [8, 9], linear programming [6] and relaxation-based methods [10]. More recently, this assignment problem handling detection errors and occlusion has been optimized using an Hungarian search algorithm [11].

In this paper, we propose an energy minimization-based approach for object trajectory reconstruction in 2D+t and 3D+t images of cellular and sub-cellular structures acquired by video-microscopy. For the correspondence problem, an object is allowed to split into two objects, merge with another object, born, die, and finally enter and leave the field. This is strongly motivated by the highly complex protein-protein interactions observed within the cell, and by the image resolution and signal to noise ratio, often leading to small vesicular structures impossible to precisely segment. Moreover, in cell biology, proteins dynamically associate and dissociate from the membrane structures to the cytosolic pool, leading to fluorescent object appearance and disappearances which is consistent with birth and death events. We propose to solve the problem using the simulated annealing algorithm because it offers the flexibility to modulate constraints according to physical situations, it does not require any particular track initialization and finally it is very fast since the number of variables used in the energy calculation is low.

2. ORGANIZING FEATURES FOR GRAPH REPRESENTATION

In this paper, an object is defined as a set of connex pixels obtained by image segmentation. We have chosen the "-trous" wavelet-based method to perform the 3D-segmentation of fluorescent structures of all individual 3D images [12]. We use several features to provide a reliable representation for object correspondence. Each object is described by its centre of mass $\mathbf{c}(O) = (c_1(O), c_2(O), c_3(O))$ where O denotes an object and c_i denote the spatial coordinates in the 3D volume, its average intensity $\bar{I}(O) = \sum_{\mathbf{x} \in O} I(\mathbf{x})$ where $I(\mathbf{x})$ in the intensity observed at pixel $\mathbf{x} \in \mathbb{R}^3$ and its sum of intensity defined as $S(O) = \sum_{\mathbf{x} \in O} I(\mathbf{x}) = |O|\bar{I}(O)$ where $|O| = \#\{\mathbf{x} \in O\}$. These features encode a global description of how the object appears. In order to deal with merge and split events, a feature composed of a set of two close objects $\{A, B\}$ is also considered. It is described by the features of the union of the two objects (in term of pixels), e.g. $\mathbf{c}(\{A, B\}) = \mathbf{c}(A \bigcup B)$. To deal with death and birth events, features of a empty set of object $\emptyset$ are also considered, except that the centre of mass is defined by: $\bar{I}(\emptyset) = 0$ and $S(\emptyset) = 0$.

3. GRAPH FORMULATION

Let F_i be a sequence of n frames (of $n_x \times n_y \times n_z$ pixels) corresponding to n time instances t_i, $i = 1...n$. Let $\mathcal{O}_i = \{O_i^1, O_i^2, \ldots, O_i^{r_i}\}$ be the set of r_i objects detected in frame F_i and $\mathcal{O} = \cup_{i=1,...,n} \mathcal{O}_i$. Note that the number r_i can be different at each frame.

We define the graph of possibility $G = (V, E)$ as a directed graph without self-loop and multi-edges. For each frame F_i, we denote $V_i = \{v(O_i^j)\}_{j=1...r_i}$ the set of vertices corresponding to each object of frame F_i and $V = \{V_1, \ldots, V_n\}$. We note E as the set of edges between two vertices of V. The edge $e \in E$ defined as $e = v(A)v(B)$ corresponds to a matching hypothesis between two objects A and B detected in two successive frames. Therefore an object A is a backward correspondence of an object B and B is a forward correspondence of A. Accordingly, the set E is defined as follows: $E = \{v(O_i^j)v(O_{i+1}^k), i = 1 \ldots n-1, j = 1 \ldots r_i, k = 1 \ldots r_{i+1} : \|\mathbf{c}(O_i^j) - \mathbf{c}(O_{i+1}^k)\|_2 < \kappa\}$, where κ is a threshold value set by the user to reduce the search space. This means that an object O detected in a frame F_i is connected to all the objects of the next frame $(i + 1)$ at a distance lower than κ. Finally, objects of the first frame have no backward correspondence and objects of the last frame have no forward correspondence.

Lets now define a collection of subgraphs $\{G_{s_1}, G_{s_2}, \ldots\}$ such that $G_{s_k} = (V, E_{s_k})$ satisfies the following constraints as illustrated in Figure 1:

- A vertex $v(O_i^j) \in V_i$ can be either connected in V_{i+1} to a *dummy* vertex ("death" event), to one vertex ("con-
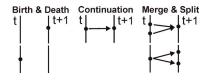

Fig. 1. Representation of the selected event types.

tinuation" event), or to two vertices ("split" event).

- A vertex $v(O_i^j) \in V_i$ can be either connected in V_{i-1} to a *dummy* vertex ("birth" event), to one vertex ("continuation" event), or to two vertices ("merge" event).

- If a vertex $v(O_i^j) \in V_i$ is linked to both $v(O_{i+1}^k)$ and $v(O_{i+1}^l)$ in V_{i+1} (split event), then $v(O_{i+1}^k)$ and $v(O_{i+1}^l)$ will only be linked in V_i to $v(O_i^j)$. Moreover, if a vertex $v(O_{i+1}^j) \in V_{i+1}$ is linked both to $v(O_i^k)$ and $v(O_i^l)$ in V_i (fusion event), then $v(O_i^k)$ and $v(O_i^l)$ will only be linked in V_{i+1} to $v(O_{i+1}^j)$. This implies for example that a forward correspondence involved in a "split" event, cannot simultaneously be involved in a "merge" event.

Finally we denote $G_{s*} = (V, E_{s*})$ as the optimal solution to the multiframe correspondence problem.

4. ENERGY FUNCTIONS

The global energy $J(E_{s_k})$ associated to the assignment problem is defined as the sum of the individual energies of all events for all times. The energy of any event can be written as $\Sigma(\mathcal{A}, \mathcal{B})$ where $\mathcal{A} \subset \mathcal{O}_i$ and $\mathcal{B} \subset \mathcal{O}_{i+1}$.

Depending on the event type, subsets $\mathcal{A}$ and $\mathcal{B}$ are defined as follow:
- "Continuation" event: $\mathcal{A} = O_i^j$ and $\mathcal{B} = O_{i+1}^k$
- "Split" event: $\mathcal{A} = O_i^j$ and $\mathcal{B} = \{O_{i+1}^k, O_{i+1}^l\}$
- "Merge" event: $\mathcal{A} = \{O_i^j, O_i^k\}$ and $\mathcal{B} = O_{i+1}^l$
- "Birth" 'event: $\mathcal{A} = O_i^j$ and $\mathcal{B} = \emptyset$
- "Death" event: $\mathcal{A} = \emptyset$ and $\mathcal{B} = O_{i+1}^j$

The energy $\Sigma(\mathcal{A}, \mathcal{B})$ is defined by:

$$\Sigma(\mathcal{A}, \mathcal{B}) = n_e \left\{\Sigma_\mathbf{c}(\mathcal{A}, \mathcal{B}) + \Sigma_{\bar{I}}(\mathcal{A}, \mathcal{B}) + \Sigma_S(\mathcal{A}, \mathcal{B})\right\} \quad (1)$$

$$\Sigma_\mathbf{c}(\mathcal{A}, \mathcal{B}) = \begin{cases} 1 & \text{for death and birth events} \\ \frac{\|\mathbf{c}(\mathcal{A}) - \mathbf{c}(\mathcal{B})\|_2^2}{\sigma_\mathbf{c}^2} & \text{otherwise} \end{cases} \quad (2)$$

$$\Sigma_{\bar{I}}(\mathcal{A}, \mathcal{B}) = \frac{(\bar{I}(\mathcal{A}) - \bar{I}(\mathcal{B}))^2}{\sigma_{\bar{I}}^2} \quad (3)$$

$$\Sigma_S(\mathcal{A}, \mathcal{B}) = \frac{(S(\mathcal{A}) - S(\mathcal{B}))^2}{\sigma_S^2} \quad (4)$$

1021

where n_e defines the number of objects implicated in the event, $n_e = 1$ for "birth" and "death" events, $n_e = 2$ for a "continuation" event and $n_e = 3$ for "merge" and "split" events. $\sigma_{\mathbf{c}}, \sigma_I$ and σ_S are scale parameters describing the standard deviations associated to each feature. In the case of birth (resp. death events), the localization energy $\Sigma_c(\emptyset, O_i^j)$ (resp. $\Sigma_c(O_i^j, \emptyset)$) is set to 1 in order to consider the subset $\emptyset$ at a distance σ_c from the object O_i^j. The definition 2 allows handling the missed detections and false alarms due to segmentation errors, a situation often encountered with noisy images. A birth event followed by a death event will be considered as a false alarm, while a death event followed by a birth event will be considered as a missed detection. In this last case, the object trajectory will be split.

5. SIMULATED ANNEALING ALGORITHM

Thus far, we are able to compute the energy associated to a solution E_{s*}. The objective of the proposed algorithm is to find the solution E_{s*} that best minimizes the global energy.

5.1. Initialization

Scale parameters $\sigma_{\mathbf{c}}, \sigma_I$ and σ_S are used to control the standard deviations associated to each feature. These values change according to the different parameters of the moving objects and the acquisition frame rate. They can be set by the user for a particular experiment or guessed by resolving the object tracking using a simple algorithm like the nearest neighbour. E_{s_1} is initialized such as each vertices have not forward and not backward correspondence.
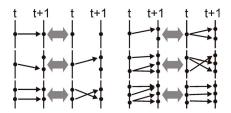
5.2. Simulated Annealing

We use the simulated annealing algorithm to reach the optimal solution E_{s*} starting from the initial solution E_{s_1} while decreasing the temperature T. At the kth iteration, a local change in the graph E_{s_k} is tested and eventually kept. All possible changes are summarized in Figure 2. The algorithm is as follows:

Initialization of the $\sigma_.$, $E_s = E_{s_1}$ and T
Do:
   An object O_i^j is randomly chosen
   Amongst all the possible changes implicating $v(O_i^j)$ and its
     neighbourhood, one is randomly chosen
   The energy variation ΔJ is computed
   The change is accepted on E_s if $\Delta J < 0$ or with a probability
     $exp(-\Delta J / T)$
   T is exponentially decreased
Until convergence
$E_{s*} = E_s$

We consider that the convergence is reached when no more change is allowed. As a change in the loop involves 1 or 2
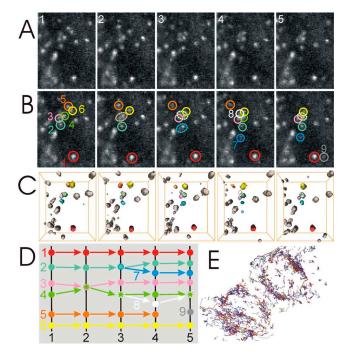

Fig. 2. Different changes of configuration operated during the optimization algorithm. The grey double arrows indicates that the left configuration can be replace be the right one. Changes of configuration implicating merge events are not represented because they are similar to changes implicating split events.

events, ΔJ requires the calculation of maximum 4 local energies Σ (2 old and 2 new events). Therefore, each iteration is performed in a very short time.

6. RESULTS

We applied our methodology to the dynamic analysis of LAR-PTP (Leukocyte common Antigen-related protein tyrosine phosphatase). Rat bladder carcinoma NBT-II cells were transitory transfected with a recombinant pcDNA3.1 plasmid containing the entire sequence of the human LAR-PTP C-terminally tagged with GFP (HA. and JJ., unpublished). 24 hours after, LAR-GFP was acquired using a "rapid" 4D deconvolution microscope as described in [13, 14], revealing highly dynamic cytoplasmic vesicles. A time series of two cells just after division (Fig. 3) was imaged at 5 3D images per seconds during 11 minutes. 130 3D images ($320 \times 350 \times 32$) pixels were collected and analysed. An average of 600 objects per image were detected by segmentation. The execution time of the tracking algorithm (C++ compiled code) was $55min$ for $\kappa = +\infty$, $70s$ for $\kappa = 20$, $45s$ for $\kappa = 10$, on a PC $2.8GHz$ with $1Go$ RAM. Examples of tracked structures in a sub region of the data showing all the event types are displayed in the Fig. 3. 3391 trajectories are obtained, including 1456 birth, 1510 death, 609 merge and 663 split events. The measured average distance between objects and the cell centre is 55 pixels ($7\mu m$) with a standard deviation of 11.5 pixels, while the membrane is located at about 70 pixels from the cell centre. This shows that vesicular structures are highly concentrated in a spherical surface, concentric to the post-mitotic cells. Moreover, the analysis of speed directionality reveals that the displacements are almost orthoradial. These observations suggests that the cytoskeleton network, responsible of the vesicle motion, is also mostly orthoradial. This is certainly due to the cell shape after mitosis.

The analysis of the global mean square displacement of all the vesicular structures shows at the short-time scale ($< 10s$) an obstructed diffusion (t^α, $\alpha = 0.75$) which slightly becomes a directed motion ($\alpha = 1.2$) for long-time scale

1022

Fig. 3. (A-B) Thumbnails showing 5 successive frames (from $t = 1$ to $t = 5$) showing several fluorescent moving structures displayed with MIP. (B) Result of tracking superimposed over MIP representation for some particular structures. Each color corresponds to a structure tracked over time and labelled from 1 to 9. Two color circles represent colors of two structures involved in a split or merge event. For example, at $t = 5$, structures 4 (green) and 8 (white) merge into the structure 4-8 (green-white) or at $t = 2$, the structure 3-4 (green-white) splits into structures 3 and 4. (C) Surface rendering of the same frames using the same colors than in (B). (D) Schematic representation of the tracking result using same numbers and colors than in (B). (E) Surface rendering of all the trajectories detected in the whole time series

($> 30s$). We can suppose that the objects are confined within small volumes moving slowly along the cytoskeleton network. This hypothesis will need to be further confirmed by stochastic simulations.

7. CONCLUSION

Thanks to the proposed tracking algorithm, we successfully investigate the dynamic of a complex membrane trafficking within living cell. It provides a powerful tool for membrane protein screening by systematic comparison of their dynamic under various conditions.

8. REFERENCES

[1] T E Fortmann, Y Bar-Shalom, and M Sheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, Jul 1983.

[2] D B Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec 1979.

[3] Auguste Genovesio and Jean-Christophe Olivo-Marin, "Split and merge data association filter for dense multi-target tracking.," in *ICPR (4)*, 2004, pp. 677–680.

[4] Ingemar J Cox, "A review of statistical data association techniques for motion correspondence," *International Journal of Computer Vision*, vol. 10, no. 1, pp. 53–66, Feb 1993.

[5] Stephane Bonneau, Maxime Dahan, and Laurent D. Cohen, "Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume.," *IEEE Transactions on Image Progressing*, vol. 14, no. 9, pp. 1384–1395, 2005.

[6] S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, 1979.

[7] I. K. Sethi and R. Jain, "Finding trajectories of feature points in a monocular image sequence," *IEEE TPAMI*, vol. 9, no. 1, pp. 56–73, 1987.

[8] Michael Jenkin, "Tracking three dimensional moving light displays (abstract only)," *SIGGRAPH Comput. Graph.*, vol. 18, no. 1, pp. 25–25, 1984.

[9] Krishnan Rangarajan and Mubarak Shah, "Establishing motion correspondence," *CVGIP: Image Underst.*, vol. 54, no. 1, pp. 56–73, 1991.

[10] Stephen T. Barnard and William B. Thompson, "Disparity analysis of images," *IEEE TPAMI*, vol. 2, no. 4, pp. 333–340, July 1980.

[11] Cor J. Veenman, Marcel J. T. Reinders, and Eric Backer, "Resolving motion correspondence for densely moving points," *IEEE TPAMI*, vol. 23, no. 1, pp. 54–72, 2001.

[12] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image processing and data analysis: the multiscale approach*, Cambridge University Press, New York, NY, USA, 1998.

[13] J B Sibarita, H Magnin, and J de Mey, "Ultra-fast 4D microscopy and high throughput distributed deconvolution," *IEEE ISBI*, pp. 769–772, 2002.

[14] Jean-Baptiste Sibarita, "Deconvolution microscopy," *Adv Biochem Eng Biotechnol*, vol. 95, pp. 201–243, 2005.