NETWORK TOMOGRAPHY FOR TRAFFICKING SIMULATION AND ANALYSIS IN FLUORESCENCE MICROSCOPY IMAGING

T. Pécot[1,2], J. Boulanger[1,2], C. Kervrann[1,2], P. Bouthemy[1]

[1]IRISA / INRIA Rennes, Campus Universitaire de Beaulieu, F-35042 Rennes
[2]INRA, UR341 Mathématiques et informatique appliquées, F-78352 Jouy-en-Josas

ABSTRACT

GFP-tagging and time-lapse fluorescence microscopy can be considered as investigation tools to observe molecular dynamics and interactions in live cells at both the microscopic and nanoscopic scales. Consequently, it is imperative to develop novel image analysis techniques able to quantify dynamics of biological processes observed in such image sequences. This motivates our present research effort which is to develop novel methods to extract information from nD data. In trafficking analysis, object tracking using conventional techniques can be very hard or impossible, especially when more than one hundred small and poorly distinguishable objects interact. However, determining the full trajectories of all the objects are not needed to monitor the cell activity. Indeed, estimating origin and destination regions of the objects of interest may be more relevant. In this paper, we propose an original approach to recover the origin and destination pairs from traffic information. Thus, we propose to consider the membrane trafficking as a road trafficking, and for the first time we exploit the recent advances in Network Tomography (NT) commonly used in network communication for biological trafficking analysis. This idea is demonstrated on realistic artificial image sequences for the Rab6 protein, a GTPase involved in the regulation of intracellular membrane trafficking.

Index Terms— Tracking, fluorescence, biomedical microscopy, biomedical signal analysis, traffic control (transportation).

1. INTRODUCTION

The Rab protein family (small GTPases - more than 60 in human cells) plays multiple roles through their association to internal membranes. Each of the members of this Rab protein family exists under different dynamic states in the cell. They exchange continuously between the cytosol and the target membrane, and are associated to transport intermediates which are routed along the cytoskeleton network. Finally, they may associate, most probably transitory, to the membrane of destination, that is the acceptor compartment. Studying the role of Rab proteins inside multiprotein complexes is then fundamental to deeply understand the molecular mechanisms responsible for membrane transport and for the maintenance of the integrity and global architecture of the cell, in space and time.

As a biological model, we will focus on the particular Rab6 protein involved in the regulation of different steps in membrane trafficking. Typically, cells expressing GFP-Rab6 show vesicles heterogeneously moving along the microtubule network, using molecular motors, from the Golgi Apparatus to the Endoplasmic Reticulum. In general, the proteins are embedded into vesicles which movements are assumed to be dependent on microtubules. Quantifying membrane trafficking that involves the movement of small transport vesicles moving along polymers (microtubules-microfilaments) from donor to acceptor compartments within the living cell is also challenging for image analysis techniques.

Actually, no existing method can be easily adapted for simultaneously and reliably tracking several hundred small similar objects with variable velocities, and for generating information on partial or complete trajectories. Stochastic temporal filtering methods that estimate trajectories of objects as they move over time [1, 2] encounter difficulties when the number of objects and clutter increases, as it is the case in trafficking analysis observed in video-microscopy. Typically, measurements from clutter and multiple objects puzzle tracking methods. It is then necessary to associate measurements with the correct object, i.e., to solve the difficult data association problem [1]. So far, data association even combined with sophisticated particle filtering techniques [2], connexionist strategies [3] or network-theory based methods [4, 5] are problematic to track several hundreds of similar objects.

A deterministic approach has also been investigated which exploits the fact that vesicles are moving along the microtubule network and thus follow the same paths [4]. The localization of paths is given by the Maximum Intensity Projection (MIP) map in the direction of time axis. The analysis of the time intensity profile on the extracted paths is achieved by using the so-called kymograms. The main drawback of the kymogram-based method is that it cannot provide a global analysis. Indeed, each path is independently supervised.

In this paper, we propose to partially solve the tracking problem which is known to be complex in trafficking analysis and biological applications. The idea is to construct and use a network that mimics membrane trafficking and aims at

estimating the origin and destination vertices of the network. This can be accomplished by only measuring the activity on the graph edges corresponding to the paths, according to the Network Tomography (NT) approach [6]. The NT-based approach, already applied in video surveillance [7, 8], would simplify the tracking process because it only requires detection of an object as it moves from one region to another, and avoids the difficult data association problem. Unlike [4, 5] where objects are tracked over time using graph matching, our idea is then to adapt the NT concept to the simulation and to the estimation of trajectories of vesicles. It just requires to count the number of vesicles that pass through each transition of the network. We will demonstrate that this method is appropriate to the investigation of membrane trafficking, in the case of vesicles moving along a filamentous network.

2. NETWORK TOMOGRAPHY

In our application, the dynamic scene is formed by vesicles moving along a dense microtubule set, that is a network of interconnecting regions of interest (see Fig. 1). This network can be described as a graph $G(E, V)$ composed of n *vertices* and r *edges*, where E denotes the set of edges, and V the set of vertices. A connection between two vertices is called a *path*, and each path consists of one or more unidirectional or bidirectional edges. In the NT-based approach, the measurements are the number of objects detected as going from one vertex to another vertex in the network. Based on these measurements, the goal is to estimate how many vesicles coming from an *origin vertex* go to a *destination vertex* along a path, in the set of all possible *origin-destination (OD) pairs* in the network, that is $c = n(n-1)$ OD pairs. This problem is then similar to determine the source-destination trafficking based on link measurements in computer networks [6]. In this approach, it is not necessary to track an object through a dynamic scene, but just to determine when an object reaches a vertex, which is generally easier than estimating a continuous trajectory.

More formally, let $x_{j,t}$, $j = 1, \cdots, c$ be the number of "transmitted" vesicles on the OD pair j at time t. The measurements $\mathbf{y}_t = (y_{1,t}, \ldots, y_{r,t})^T$ are the number of vesicles that pass from one vertex to another vertex at time t. The inherent randomness of the measurements motivates the adoption of a statistical approach. Typically, we reasonably assume that the traffic is temporally distributed as a Poisson process, $x_{j,t} \sim \text{Poisson}(\lambda_j)$. In this traffic flow problem, we then assume the following model $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, where $\mathbf{x}_t = (x_{1,t}, \ldots, x_{c,t})^T$, and $\mathbf{A}$ denotes a $r \times c$ routing matrix which binary elements $a_{ij} = 1$ if edge i is in the path for the OD pair j, and 0 otherwise.

Numerically, the number c is greater than r, and the problem is then *ill-posed*. Additional constraints are necessary for solving the *inverse problem*. Vardi [6] proposed to introduce constraints related to the assumption that the traffic is temporally Poisson distributed. The NT method amounts then to estimating the values λ_j given the additional set of equations corresponding to temporal averages:

$$
\begin{cases}
\overline{y_i} = \sum_{k=1}^{c} a_{i,k}\lambda_k, & i = 1, \ldots, r, \\
cov(y_i, y_{i'}) = \sum_{k=1}^{c} a_{i,k}a_{i',k}\lambda_k, & 1 \le i \le i' \le r.
\end{cases}
$$

This set of equations gives a system of $r(r+3)/2$ linear equations that forms an over-constrained problem that can be better solved with the conditions $\lambda_i \ge 0$. Moreover, in this application, the aim is not to obtain the number of vesicles that utilize each path, but to estimate the proportions of vesicles on each path. Hence, unlike previous methods [6, 7, 8], we impose the condition $\sum_{i=1}^{c} \lambda_i = 1$ as an additional constraint, that will greatly improve the results. The previous system can be written more compactly as

$$
\begin{pmatrix} \overline{\mathbf{y}} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{\Lambda}, \tag{1}
$$

where $\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_c)^T$ contains the mean of the traffic flow, $\mathbf{s} = \{cov(y_i, y_{i'})\}$ is the sample covariance matrix rewritten as a vector of length $r(r+1)/2$, and $\mathbf{B}$ is an $(r(r+1)/2) \times c$ matrix with the (i, i')th row of $\mathbf{B}$ being the element-wise product of row i and row i' of the matrix $\mathbf{A}$.

The system can be solved using the estimation-maximization (EM) method [6, 7] or the convex-projection algorithms [8]. In our case, we adapt a non negative mean square estimation which also provides a simple and reliable way to estimate the origin-destination traffic $\widehat{\mathbf{\Lambda}}$. In the implementation, our method is then based on the *lsqnonlin* function from the Matlab Optimization toolbox. Note that a review of existing methods is also proposed in [9].

In order to demonstrate the performance of this approach, we have designed realistic simulations. In the next section, we describe the "generator" used for these simulations of fluorescence image sequences based on the NT framework.

3. NETWORK TOMOGRAPHY-BASED IMAGE SEQUENCE MODELING

For the applicability and performance evaluation of the NT-based approach for trafficking analysis, we designed a generator able to mimic dynamics related to the GFP-Rab6 expression [10, 11]. Typically, the moving tagged vesicles observed with fast 4D deconvolution microscopy (wide-field) [4] look like small bright blobs in real image sequences. The vesicle diameters (from 60 nm to 150 nm) are often below the resolution of the microscope (about $130 \times 130 \times 300$ nm). However, the point spread function of the video-microscope makes them appear as larger structures even if a deconvolution process [12] is applied. Furthermore, when the object density increases, vesicles gather together and constitute small rods.

The vesicles move along microtubules propelled by two classes of molecular motors (dynein and kinein proteins), associated to microtubules. It has been shown that the concentration of these molecular motors influences the structure and the dynamics of the microtubule network. In stable conditions, the regime of these motors is assumed to be constant. This explains why the observed velocity of vesicles seems constant if they move along the same microtubule. The generator we propose is based on this description and on the NT concept explained in the previous section.

3.1. Models for geometry and appearance

In 2D image sequences expressing GFP-Rab6, the vesicles can be represented by 2D anisotropic Gaussian blobs with variances related to the blob size ranging from 60 nm to 150 nm, that is close to the pixel size. The major axis of the ellipse is aligned with the direction of the movement of vesicles, also given by the spatial orientation of microtubules. We impose that the anisotropy is higher for vesicles with high velocities. This modeling based on oriented anisotropic Gaussian blobs is able to describe the main dynamical components seen in real images, included small rods with variable lengths corresponding to "trains" of nearby vesicles moving along the same microtubule, and probably with the same velocity.

3.2. NT-based model for dynamics

Network modeling In order to generate synthetic but realistic microtubule networks, we exploit real fluorescence image sequences showing only moving tagged vesicles. In general, direct GFP-tagged microtubule networks are very complex for image analysis methods and experts, and individual microtubules cannot be easily extracted. However, a set of significant microtubules in the network can be computed from a maximum intensity projection (MIP) map in the direction of time axis. Fig. 1 contains the MIP map obtained from a sequence of 300 images. The bright paths are generated by tagged vesicles from frame to frame, and give an insight of the microtubule network under concern.

However, as shown in Fig. 1, all the paths are not completely defined in the MIP maps, especially if the sequence duration is too short. In our application, the gaps are manually completed by using a "painting" software. Furthermore, the locations of the selected curves are estimated using the unbiased line detection algorithm [13] applied to the MIP map, and are called *roads*. Finally, each road is defined by a list of points, and each point of this list is parameterized by the width of the road, and its orientation (see Fig. 1). From the set of roads, the intersections are easily found and defined as vertices. Furthermore, each road given by the line detection algorithm is associated to a couple of edges and the bi-directional graph $G(E, V)$ is then fully defined.

Selection of origin/destination vertices In the proposed simulation, once the network has been computed, the biologist
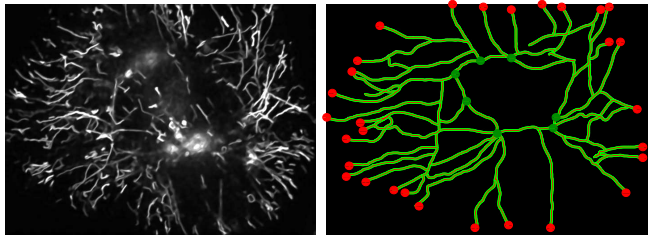

Fig. 1. Representation of a reasonable synthetic network from the MIP map. Left: MIP map computed from an image sequence. The paths used by the vesicles appear as bright filaments; right: the network after manual simplifications, is composed of 146 vertices and 160 bi-directional roads, which correspond to 320 directional edges in the graph associated to the network. The origin vertices are labeled in green, while the destination vertices are labeled in red.

has to specify the origin (e.g., donor organelle) and destination (e.g., acceptor organelle) vertices on the network, that corresponds to the OD pairs $\mathbf{\Lambda}$ in Section 2. The other vertices represent the intersection points of the network and are only used for the routing. These labels are assumed to be biologically motivated locations, like the Golgi Apparatus or some endpoints in the case of the Rab6 analysis [14].

In Fig. 1, the origin vertices corresponding to the membrane of the Golgi apparatus are manually selected and labeled in green, while the destination vertices labeled in red correspond to endpoints. In this example, vesicles are imposed only to go from the Golgi to the endpoints located at the periphery of the cell.

In our approach, we assume that the paths are defined as the *minimal paths* between the origin and the destination vertices. They are computed using the Dijkstra [15] algorithm on the graph $G(E, V)$, that yields the routing matrix $\mathbf{A}$. In that case, the weight associated to each edge of E can be defined as a function of the length of the corresponding road, but other features (e.g., velocity, congestion in trafficking) could be also used. During the simulation, an OD pair is selected for each vesicle according to a multinomial law of parameters $\mathbf{\Lambda}$.

Finally, the obtained simulation show vesicles moving along the estimated roads with a distance proportional to the velocity associated to the roads, following the shortest path between origins and destinations.

4. EXPERIMENTAL RESULTS

We propose two experiments to analyze the NT estimation method. They are reported in Fig. 2. In the first row, we have displayed the networks with the origin vertices labeled in green and the destination vertices labeled in red. The second row shows a frame of the corresponding simulated image sequences where we can distinguish the vesicles appearing as white blobs. Finally, in the last row, the blue circles correspond to the results of the estimation procedure obtained with the *lsqnonlin* function, while the red crosses represent the ground truth $\mathbf{\Lambda}$.

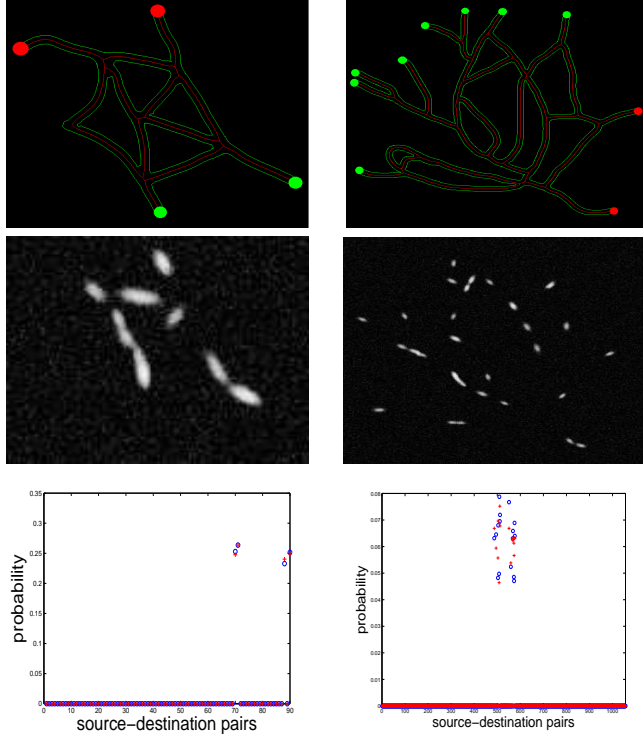In the first experiment, we have considered a simple star-

Fig. 2. Experiments. Fisrt column: star-shaped network with $n = 10$ vertices, $p = 90$ OD pairs. Simulation of 1000 images containing 10 moving vesicles. Second column: network that mimics real data with $n = 23$ vertices, $p = 506$ OD pairs. Simulation of 1000 images containing 75 moving vesicles. First row: networks with origin vertices in green, and destination vertices in red. Second row: frames of the corresponding simulated image sequences. Third row: estimation of the traffic flow with the NT-based approach (blue circles), and ground truth (red crosses).

shaped network with $n = 10$ vertices, and 4 OD pairs were selected among the $p = 90$ possible OD pairs. A sequence of 1000 images 126×100 pixels containing 10 moving vesicles was simulated. We can see on the right column in Fig. 2 that the estimation $\hat{\Lambda}$ is close to the ground truth (MSE $= E[(\Lambda - \widehat{\Lambda})^2] \simeq 10^{-6}$).

The second experiment is more complex, and the network is composed of $n = 23$ vertices. In this experiment, a sequence of 1000 images 360×455 pixels containing 75 vesicles was simulated, and 16 origin-destination pairs were selected among $p = 506$. In that case, it has to be noted that the estimated OD pairs correspond to the OD pairs effectively followed by the vesicles, even if the proportions of vesicles on each path are not accurately estimated (MSE $\simeq 6.2 \times 10^{-7}$).

5. CONCLUSION

We have demonstrated the potential of Network Tomography for trafficking simulation and analysis in fluorescence microscopy imaging. The applicability of NT-based approach is two-fold: *i*) the concept has been used to generate dynamics in fluorescence image sequences that mimic real data where particles are moving along microtubule networks. If well designed, the NT-based simulation is also a way to define a com-

putational model that can explain trafficking; *ii*) the estimation scheme allows us to infer information about the global status of the traffic from local count measurements. The data association problem which is too difficult to be solved by tracking algorithms is then potentially alleviated. The advantage of testing the NT-based approach on simulations is that a certain number of features can be controlled like the number of vesicles moving on the microtubule network, their velocities, or the graph topology. It is worth noting that this framework can be used to evaluate the performance of already developped object detection techniques as usually used in video-microscopy. For future research, it will be necessary to find a fully automatic technique to extract the network from real sequences for high-throughput experiments, and to evaluate the robustness of the NT estimation method with respect to object detection errors when applied to real image sequences.

6. REFERENCES

[1] A. Genovesio, T. Liedl, V. Emiliani, W.J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin, "Multiple particle tracking in 3D+t microscopy: Method and application to the tracking of endocytosed quantum dots," *IEEE Trans. on IP*, vol. 15, no. 5, pp. 1062–1070, 2006.

[2] I. Smal, W. Niessen, and E. Meijering, "Bayesian tracking for fluorescence microscopic imaging," in *Proc. of IEEE ISBI'2006*, Washington, Apr. 2006, pp. 550–553.

[3] V. Racine, A. Hertzog, J. Jouaneau, J. Salamero, C. Kervrann, and J.B. Sibarita, "Mutliple target tracking of 3d fluorescent objects based on simulated annealing," in *Proc. of IEEE ISBI'2006*, Apr. 2006.

[4] J.B. Sibarita, V. Racine, and J. Salamero, "Quantification of membrane trafficking on a 3d cytoskeleton network in living cells," in *Proc. of IEEE ISBI'2006*, Apr. 2006.

[5] D. Thomann, J. Dorn, P.K. Sorger, and G. Danuser, "Automatic fluorescent tag localization ii: improvement in super-resolution by relative tracking," *Journal of Microscopy*, vol. 211, no. 3, pp. 230–248, Sept. 2003.

[6] Y. Vardi, "Network tomography: Estimation source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, Mar. 1996.

[7] S. Santini, "Analysis of traffic flow in urban areas using web cameras," in *Fifth IEEE Workshop on Applications of Computer Vision*, 2000, vol. 596, pp. 140–145.

[8] J. E. Boyd, J. Meloche, and Y. Vardi, "Statistical tracking in video traffic surveillance," in *Proc. of ICCV'99*, 1999, vol. 1, pp. 163–168.

[9] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: existing techniques and new directions," in *SIGCOMM '02: Proc. Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications*, New York, USA, 2002, pp. 161–174.

[10] J. Boulanger, C. Kervrann, and P. Bouthemy, "Simulation and estimation of fluorescence in microscopy image sequence," in *Proc. of the MICCAI Workshop on Microscopic Image Analysis and Application to Biology*, Copenhague, Oct. 2006.

[11] J. Boulanger, C. Kervrann, and P. Bouthemy, "Estimation of dynamic background for fluorescence video-microscopy," in *Proc. of ICIP'2006*, Atlanta, Oct. 2006.

[12] J.-B. Sibarita, H. Magnin, and J.R. De Mey, "Ultra-fast 4D microscopy and high throughput distributed deconvolution," in *Proc. of IEEE ISBI'2002*, Washington, June 2002, pp. 769–772.

[13] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. on PAMI*, vol. 20, no. 2, pp. 113–125, 1998.

[14] J. White, L. Johannes, F. Mallar, A. Girod, G. Stephan, S. Reinsh, P. Keller, B. Tzschaschel, A. Echard, B. Goud, and E. Stelzer, "Rab6 coordinates a novel golgi to ER retrograde transport pathway in live cells," *The Journal of Cell Biology*, vol. 147, no. 4, pp. 743–760, Nov. 1999.

[15] E.W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, Dec. 1959.