

Tomographie de réseau appliquée à la simulation et à l'analyse de trafic dans des séquences d'images de microscopie à fluorescence

T. Pécot^(1,2), J. Boulanger^(1,2), C. Kervrann^(1,2), P. Bouthemy⁽¹⁾

⁽¹⁾IRISA / INRIA Rennes, Campus Universitaire de Beaulieu, 35042 Rennes

⁽²⁾INRA, UR341 Mathématiques et informatique appliquées, 78352 Jouy-en-Josas

Thierry.Pecot@irisa.fr, Charles.Kervrann@irisa.fr, Jerome.Boulanger@irisa.fr, Patrick.Bouthemy@irisa.fr

Résumé – La technique de marquage avec la protéine GFP “Green Fluorescent Protein” et la vidéomicroscopie à fluorescence sont des outils d’investigation permettant d’observer des dynamiques et des interactions moléculaires dans des cellules vivantes, tant à l’échelle microscopique qu’à l’échelle nanoscopique. Par conséquent, il est impératif de développer de nouvelles techniques d’analyse d’images capables de quantifier les dynamiques des processus biologiques observés dans ces séquences. Ceci motive notre effort de recherche qui consiste à développer de nouvelles méthodes d’extraction d’informations à partir de données nD. Dans l’analyse de trafic, le suivi d’objets s’appuyant sur des techniques conventionnelles peut s’avérer très complexe, voire impossible, surtout quand un grand nombre de petits objets coalescents sont en interaction. Néanmoins, l’estimation des trajectoires complètes de tous les objets n’est pas toujours nécessaire à la compréhension et la mesure de l’activité cellulaire. En effet, estimer les régions “origine” et “destination” de ces objets peut s’avérer plus pertinente. Dans cet article, nous proposons une approche originale pour inférer les zones “origine” et “destination” à partir d’informations partielles relatives au trafic. Ainsi, le trafic membranaire est assimilé à un trafic routier, ce qui permet alors d’exploiter les récentes avancées en Tomographie de Réseau (TR) bien connues dans la communauté réseaux de communication pour étudier le trafic vésiculaire. Cette approche est validée sur des séquences d’images artificielles relatives à la protéine Rab6, une GTPase impliquée dans la régulation du trafic membranaire intracellulaire.

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 the “origin” and “destination” regions of the objects 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.

1 Introduction

La famille de protéines Rab (GTPases - plus de 60 membres chez l’homme) joue de multiples rôles lors de son association avec les membranes internes. Chacun des membres de cette famille de protéines existe sous différents états dynamiques dans la cellule : *i*) diffusion dans le cytosol; *ii*) échange entre le cytosol et les membranes; *iii*) transport vésiculaire. L’étude du rôle des protéines Rab au sein des complexes multiprotéiques est fondamentale pour aider à la compréhension des mécanismes moléculaires responsables du transport membranaire et du maintien de l’intégrité et de l’architecture globale de la cellule, dans l’espace et le temps.

Dans notre étude, nous nous intéressons plus particulièrement à la protéine Rab6 impliquée dans la régulation de différentes étapes intervenant dans le trafic membranaire. Typiquement, les cellules qui expriment la protéine chimère GFP-Rab6 mettent en évidence des vésicules (intermédiaires de transport) qui se déplacent de manière hétérogène le long du réseau de microtubules et qui sont propulsées par des moteurs moléculaires, depuis l’appareil de Golgi jusqu’au Reticulum Endoplasmique.

Actuellement, aucune méthode existante n’est capable de suivre simultanément et fiablement plusieurs centaines d’objets

similaires affublés de vitesses différentes. Les méthodes de filtrage stochastique temporel qui estiment les trajectoires des objets au cours du temps [1, 2] éprouvent des difficultés quand le nombre d’objets est important, comme c’est le cas dans l’analyse du trafic en vidéomicroscopie. En effet, lorsque chaque objet est suivi, il est nécessaire d’associer les mesures aux objets correspondants image par image, et ainsi, de résoudre le difficile problème d’association de données [1]. Jusqu’ici, l’association de données, même combinée à des techniques sophistiquées de filtrage particulière [2], de stratégies connexionistes [3] ou de méthodes fondées sur des mises en correspondance de graphes [4, 5] ne parvient pas à suivre plusieurs centaines d’objets similaires avec un taux d’erreurs faible.

Des approches déterministiques ont aussi été développées. Dans [4], les auteurs exploitent le fait que les vésicules se déplacent le long du réseau de microtubules et suivent donc le même chemin. La localisation des chemins est obtenue à partir de la carte de Projection du Maximum d’Intensité (PMI) dans la direction temporelle, sur toute la séquence d’images. L’analyse temporelle du profil d’intensité sur les chemins extraits est réalisée en utilisant des “kymogrammes”. La principale limitation de cette méthode est qu’elle ne peut produire une analyse globale du trafic observé. En effet, chaque chemin est traité indépendamment.

Nous proposons de résoudre partiellement le problème de suivi qui est connu pour être complexe dans l'analyse de trafic et les applications biologiques. L'idée est d'exploiter un réseau sur lequel repose le trafic membranaire, et d'estimer les origines et destinations des intermédiaires de transport. Ceci peut être accompli en mesurant l'activité observée sur chacune des arêtes du graphe associé au réseau, selon une approche par la Tomographie de Réseau (TR) [6]. L'approche s'appuyant sur la TR, déjà expérimentée en vidéosurveillance [7, 8], simplifie le processus de suivi car il suffit de détecter les objets qui se déplacent d'une région à l'autre, et évite ainsi le problème d'association de données. A l'inverse des approches [4, 5] où les objets sont suivis au cours du temps en effectuant des mises en correspondance de graphes, notre idée est d'adapter le concept de la TR à la simulation et à l'estimation des trajectoires de vésicules. Il suffit alors de compter le nombre de vésicules qui franchissent chaque arête du réseau. Nous démontrerons la validité de cette méthode pour faciliter la compréhension du transport membranaire dans le cas où les vésicules se déplacent le long d'un réseau de microtubules.

2 Tomographie de réseau

Dans notre problématique, la scène dynamique correspond à des vésicules se déplaçant le long d'un réseau dense de microtubules, c'est-à-dire un réseau de régions d'intérêt interconnectées (voir figure 1). Ce réseau peut être décrit par un graphe $G(E, V)$ composé de n sommets et de r arêtes, où E désigne l'ensemble des arêtes, et V l'ensemble des sommets. La connexion d'un sommet à un autre est appelée *chemin*, et chaque chemin est composé d'une ou plusieurs arêtes unidirectionnelles. Dans l'approche s'appuyant sur la TR, les mesures sont définies par le nombre d'objets détectés se déplaçant d'un sommet à un autre dans le réseau. Ces mesures permettent d'évaluer le nombre de vésicules initialement localisées à un *sommet origine* et arrivant à un *sommet destination* en parcourant un chemin, parmi toutes les *paires origine-destination (OD)* possibles du réseau, c'est-à-dire $c = n(n-1)$ paires OD. Ce problème est similaire à la détermination du trafic source-destination s'appuyant sur les mesures de trafic effectuées sur les connexions de réseaux d'ordinateurs [6]. Dans cette approche, il n'est pas nécessaire de suivre un objet dans une scène dynamique, mais seulement de déterminer le moment où un objet atteint un sommet, ce qui semble plus aisé a priori que d'estimer une trajectoire continue.

Plus formellement, soit $x_{j,t}$, $j = 1, \dots, c$ le nombre de vésicules "transmises" sur la paire OD j à l'instant t . Les mesures $\mathbf{y}_t = (y_{1,t}, \dots, y_{r,t})^T$ représentent le nombre de vésicules qui transitent d'un sommet à un autre à l'instant t . Le caractère aléatoire intrinsèque des mesures motive l'adoption d'une approche statistique. Typiquement, nous pouvons raisonnablement supposer que le trafic suit temporellement une loi de Poisson, $x_{j,t} \sim \text{Poisson}(\lambda_j)$. Ensuite, nous modélisons ce problème d'estimation de flux de trafic par $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, où $\mathbf{x}_t = (x_{1,t}, \dots, x_{c,t})^T$, et \mathbf{A} représente la matrice de routage $r \times c$ dont les éléments binaires $a_{ij} = 1$ si l'arête i appartient au chemin de la paire OD j , et 0 sinon.

Le nombre c est plus grand que r et le problème est donc *mal posé*. Des contraintes additionnelles deviennent néces-

saires pour résoudre ce *problème inverse*. Vardi [6] a proposé d'introduire des contraintes relatives à l'hypothèse que le trafic est distribué temporellement suivant une loi de Poisson. La méthode de la TR parvient alors à estimer les valeurs λ_j étant données les contraintes supplémentaires correspondant aux moyennes temporelles :

$$\begin{cases} \bar{y}_i = \sum_{k=1}^c a_{i,k} \lambda_k, & i = 1, \dots, r, \\ \text{cov}(y_i, y_{i'}) = \sum_{k=1}^c a_{i,k} a_{i',k} \lambda_k, & 1 \leq i \leq i' \leq r. \end{cases}$$

Cet ensemble d'équations constitue un système de $r(r+3)/2$ équations linéaires pour lequel on impose $\lambda_i \geq 0$. En outre, dans notre problématique, le but n'est pas d'obtenir le nombre exact de vésicules qui empruntent chacun des chemins, mais d'estimer les proportions de vésicules sur chaque chemin. Ainsi, contrairement aux méthodes précédentes [6, 7, 8], nous imposons aussi la condition $\sum_{i=1}^c \lambda_i = 1$ comme une contrainte supplémentaire, ce qui améliore considérablement nos résultats expérimentaux. Le système précédent peut être réécrit plus synthétiquement comme

$$\begin{pmatrix} \bar{\mathbf{y}} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\Lambda}, \quad (1)$$

où $\boldsymbol{\Lambda} = (\lambda_1, \dots, \lambda_c)^T$ représente la moyenne du flux de trafic, $\mathbf{s} = \{\text{cov}(y_i, y_{i'})\}$ est la matrice de covariance échantillonnée réécrite comme un vecteur de longueur $r(r+1)/2$, et \mathbf{B} une matrice $(r(r+1)/2) \times c$ où la (i, i') ème ligne de \mathbf{B} correspond au produit terme à terme de la ligne i et de la ligne i' de la matrice \mathbf{A} .

Ce système peut être résolu en utilisant la méthode (EM) Espérance-Maximisation [6, 7] ou l'algorithme de projections convexes [8]. Dans notre cas, nous adoptons une estimation par la méthode des moindres carrés itérés avec une contrainte de positivité, ce qui assure une manière simple et fiable d'estimer le trafic origine-destination $\hat{\boldsymbol{\Lambda}}$. Pour mettre en oeuvre cette méthode, nous avons utilisé la fonction *lsqnonneg* de la boîte à outils de Matlab Optimization. Par ailleurs, il existe une revue des méthodes existantes dans [9].

Afin de démontrer les performances de cette approche, nous avons élaboré des simulations suffisamment réalistes. Dans le paragraphe 3, nous décrivons le "générateur" utilisé pour les simulations de séquences d'images de fluorescence s'appuyant sur le concept de la TR.

3 Modélisation de séquences d'images fondé sur la tomographie de réseau

Afin d'évaluer les performances de l'analyse du trafic fondée sur la TR, nous avons élaboré un générateur capable de reproduire des dynamiques relatives à l'expression de GFP-Rab6 [10, 11]. Typiquement, les vésicules marquées observées par microscopie à déconvolution rapide 4D (champ large) [12] ressemblent à de petits spots lumineux dans les séquences d'images réelles. Le diamètre des vésicules (de 60 nm à 150 nm) est alors souvent en dessous de la résolution du microscope (environ $130 \times 130 \times 300$ nm). La fonction de transfert optique du vidéomicroscope laisse paraître des structures vésiculaires,

même après un processus de déconvolution [12]. Enfin, lorsque la densité d'objets augmente, les vésicules se rassemblent et forment des "trains", c'est-à-dire des bâtonnets orientés.

Les vésicules se déplacent le long de microtubules propulsées par deux classes de moteurs moléculaires (dynéine et kynéine), associées aux microtubules. Il a été montré que la concentration de ces moteurs moléculaires influençait la structure et les dynamiques du réseau de microtubules. Dans des conditions stables, le régime de ces moteurs est supposé constant. Ceci explique partiellement pourquoi les vitesses observées des vésicules semblent également constantes lorsqu'elles se déplacent le long du même microtubule. Le générateur que nous proposons s'appuie sur cette description et sur le concept de la TR présenté dans le paragraphe 2.

3.1 Modèles de géométrie et d'apparence

Dans les séquences d'images 2D de cellules exprimant GFP-Rab6, les structures vésiculaires peuvent être représentées par des spots gaussiens anisotropes de variance relative à la taille du spot. Le grand axe de l'ellipse est aligné avec la direction du mouvement de la vésicule, également donné par l'orientation spatiale des microtubules. Nous imposons une anisotropie plus forte pour les vésicules dont la vitesse est plus importante. Cette modélisation reposant sur des spots gaussiens anisotropes orientés est capable de décrire les composantes dynamiques principales observées dans des séquences réelles, ainsi que les bâtonnets de tailles variables correspondant à des "trains" de vésicules proches, se déplaçant sur le même microtubule, avec une vitesse identique.

3.2 Modèle dynamique s'appuyant sur la TR

Modélisation d'un réseau Dans le but de générer un réseau de microtubules synthétique mais réaliste, nous exploitons des séquences d'images de microscopie à fluorescence réelles ne présentant que des vésicules marquées en mouvement. Un ensemble significatif de microtubules du réseau peut être obtenu à partir de la carte de projection du maximum d'intensité (PMI) dans la direction temporelle. La figure 1 présente la carte PMI obtenue à partir d'une séquence de 300 images. Les filaments lumineux sont générés par des vésicules marquées image après image, et donnent un aperçu du réseau de microtubules considéré. Les discontinuités sont traitées manuellement à l'aide d'un logiciel de dessin. Un algorithme de détection non biaisé de lignes [13] sélectionne des courbes correspondant aux routes. Chaque route est définie par une liste de points, chaque point est caractérisé par la largeur de la route et son orientation. Les intersections des routes définissent les sommets, et chaque route est associée à un couple d'arêtes. En procédant ainsi, le graphe $G(E, V)$ est complètement défini.

Sélection des sommets origine/destination Dans la simulation proposée, une fois le réseau connu, le biologiste doit spécifier les sommets origine (e.g., les organelles donneurs) et les sommets destination (e.g., les organelles accepteurs), ce qui forme les paires OD Λ du paragraphe 2. Les autres sommets représentent les points d'intersection du réseau et ne sont utilisés que pour le routage. Ces sommets origine et destination représentent des zones d'intérêt sur le plan biologique, comme

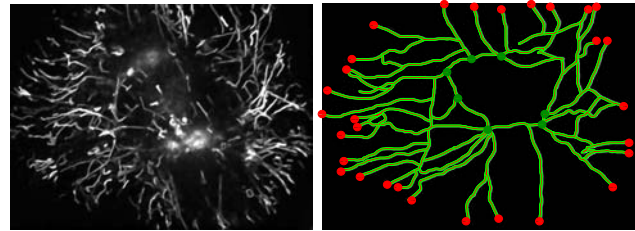


Figure 1: Représentation d'un réseau synthétique réaliste obtenu à partir de la carte PMI. Gauche : carte PMI extraite d'une séquence d'images. Les chemins empruntés par les vésicules apparaissent comme des filaments lumineux; droite : le réseau après simplifications manuelles, composé de 146 sommets et de 160 routes bidirectionnelles, ce qui correspond à 320 arêtes dirigées dans le graphe associé au réseau. Les sommets origine sont labelisés en vert, tandis que les sommets destination sont labelisés en rouge.

l'appareil de Golgi ou certains points d'arrivée dans le cas de l'analyse de Rab6 [14]. Sur la figure 1, les sommets origine correspondent à la membrane de l'appareil de Golgi et sont manuellement labelisés en vert, tandis que les sommets destination labelisés en rouge correspondent aux points d'arrivée.

Dans notre approche, nous supposons que les chemins correspondent aux *plus courts chemins* joignant les sommets origine et les sommets destination. Ils sont obtenus grâce à l'algorithme de Dijkstra [15] et au graphe $G(E, V)$, établissant la matrice de routage \mathbf{A} . Dans ce cas, les poids associés à chaque arête de E sont fonction de la longueur de la route correspondante, mais aussi d'autres paramètres (vitesse, nombre de vésicules présentes sur l'arête). Durant la simulation, une paire OD est tirée pour chaque vésicule suivant une loi multinomiale de paramètres Λ .

Finalement, la simulation obtenue présente des vésicules se déplaçant le long des courbes estimées, en parcourant à chaque intervalle de temps une distance proportionnelle à la vitesse associée aux routes, suivant le chemin le plus court entre les origines et les destinations.

4 Résultats expérimentaux

Nous proposons deux expériences pour analyser la méthode d'estimation de la TR. Elles sont illustrées sur la figure 2. Sur la première ligne de cette figure, les réseaux sont représentés avec les sommets origine en vert et les sommets destination en rouge. Sur la seconde ligne, on présente une image de chacune des séquences simulées sur lesquelles les vésicules apparaissent comme des spots lumineux. Enfin, les résultats sont illustrés sur la dernière ligne, avec des cercles bleus qui correspondent aux estimations obtenues avec la fonction $lsqnonneg$, et des croix rouges pour la vérité terrain Λ .

Dans la première expérience, nous avons considéré un réseau simple en forme d'étoile constitué de 10 sommets. Quatre paires OD ont été sélectionnées parmi les $p = 90$ paires OD possibles. Une séquence de 1000 images 126×100 pixels présentant 10 vésicules en mouvement sur chaque image a été simulée. Nous pouvons voir sur la colonne de droite de la figure 2 que l'estimation $\hat{\Lambda}$ est proche de la vérité terrain ($MSE = E[(\Lambda - \hat{\Lambda})^2] \simeq 10^{-6}$).

Le réseau associé à la seconde expérience est composé de

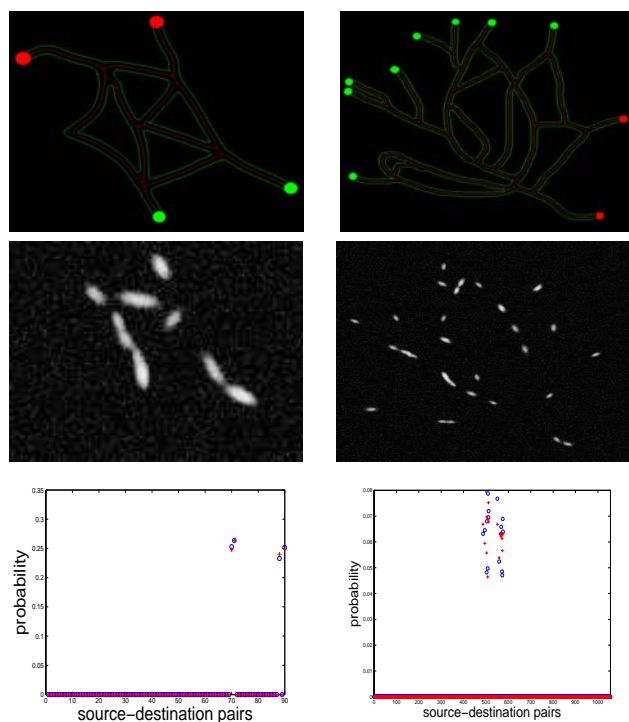


Figure 2: Expériences. Première colonne : réseau en forme d'étoile avec 10 sommets, $p = 90$ paires OD. Simulation de 1000 images contenant 10 vésicules en mouvement. Seconde colonne : réseau réaliste avec 23 sommets, $p = 506$ paires OD. Simulation de 1000 images contenant 75 vésicules en mouvement. Première ligne : réseaux avec les sommets origine en vert, et les sommets destination en rouge. Seconde ligne : images extraites des séquences simulées correspondantes. Troisième ligne : estimation du trafic avec l'approche fondée sur la TR (cercles bleus), et vérité terrain (croix rouges).

$n = 23$ sommets et possède une topologie plus complexe. Dans cette expérience, une séquence de 1000 images 360×455 pixels présentant 75 vésicules a été simulée, et 16 paires OD ont été sélectionnées parmi les $p = 506$ possibles. Dans ce cas, il est à noter que les paires OD estimées correspondent aux paires OD effectivement empruntées par les vésicules, même si les proportions estimées de vésicules sur chacun des chemins ne sont pas tout à fait exactes ($MSE \simeq 6.2 \times 10^{-7}$).

5 Conclusion

Nous avons démontré le potentiel de la tomographie de réseau pour la simulation et l'analyse du trafic dans les séquences de microscopie à fluorescence. L'intérêt de l'approche s'appuyant sur la TR est double : *i*) le concept a été utilisé pour générer les dynamiques dans des séquences d'images à fluorescence qui miment des données réelles dans lesquelles des particules se déplacent le long de réseaux de microtubules. La simulation fondée sur la TR est aussi une manière de définir un modèle qui peut aider à expliquer le trafic; *ii*) l'estimation nous permet d'inférer le fonctionnement global à partir de comptages locaux. Le problème d'association de données trop complexe pour être réglé par des algorithmes de suivi est ainsi potentiellement écarté. Le fait de tester l'approche s'appuyant sur la TR sur des simulations présente l'avantage de contrôler un certain

nombre de paramètres, comme le nombre de vésicules se déplaçant le long du réseau de microtubules, leur vitesse, ou la forme du réseau. Cet outil peut servir également à évaluer les performances de techniques de détection ou de suivi d'objets déjà développées et utilisées en vidéomicroscopie. Concernant les futures orientations, il sera nécessaire de déterminer une technique complètement automatique pour extraire un réseau depuis une séquence réelle afin de pouvoir appliquer l'estimation à un grand nombre d'expérimentations, ainsi que d'évaluer la robustesse de l'estimation obtenue par la méthode de la TR quand elle est appliquée à des séquences réelles.

References

- [1] A. Genovesio, T. Liedl, V. Emiliani, W.J. Parak, M. Coppey-Moisand, 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, "Multiple 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*, Copenhagen, 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.