HAL
open science

A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery

Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy

To cite this version:

Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy. A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery. Medical Image Analysis, 2008, 13, 132-142Video-microscopy Fluorescence Cellular biology Image sequence modeling Simulation Estimation. hal-00794917

HAL Id: hal-00794917

https://inria.hal.science/hal-00794917v1

Submitted on 13 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Simulation and Estimation Framework for Intracellular Dynamics and Trafficking in Video-microscopy and Fluorescence Imagery

Jérôme Boulanger [a,b,*] Charles Kervrann [a,b] Patrick Bouthemy [a]

[a] *IRISA/INRIA Rennes, Campus Universitaire de Beaulieu, F-35042 Rennes*

[b] *INRA, UR341 Mathématiques et Informatique Appliquées, F-78352 Jouy-en-Josas*

Special Issue MICCAI '06

Abstract

Image sequence analysis in video-microscopy has now gained importance since molecular biology is presently having a profound impact on the way research is being conducted in medicine. However, image processing techniques that are currently used for modeling intracellular dynamics, are still relatively crude and yield imprecise results. Indeed, complex interactions between a large number of small moving particles in a complex scene cannot be easily modeled, limiting the performance of object detection and tracking algorithms. This motivates our present research effort which is to develop a general estimation/simulation framework able to produce image sequences showing small moving spots in interaction, with variable velocities, and corresponding to intracellular dynamics and trafficking in biology. It is now well established that spot/object trajectories can play a role in the analysis of living cell dynamics and simulating realistic image sequences is then of major importance. We demonstrate the potential of the proposed simulation/estimation framework in experiments, and show that this approach can be also used to evaluate the performance of object detection/tracking algorithms in video-microscopy and fluorescence imagery.

Key words: video-microscopy, fluorescence, cellular biology, image sequence modeling, simulation, estimation, object detection.

* Corresponding author. Address: UMR 144 Institut Curie/CNRS, 26 rue d'Ulm, 75248 Paris cedex 05, France. Phone: +33 1 42 34 63 21 Fax: +33 1 42 43 63 30
Email address: `jerome.boulanger@curie.fr` (Jérôme Boulanger).

1 Introduction

1.1 Context in biology

The development of *systems biology* is characterized by the settlement of new techniques and technologies producing a vast amount of data of different types or origins. Only automatic approaches for analysis and interpretation of complex and massive data will allow researchers to face this new challenge. This is already well established for a number of biological fields such as DNA sequence analysis, expression data analysis, DNA micro-arrays analysis. Also, in dynamical imaging of biological samples substantial amount of work is necessary to overcome conceptual and technological obstacles. This motivates our present research effort which is to develop novel approaches based on recent methods in computer vision and signal processing, able to analyze information from 4D data related to intracellular dynamics and membrane transport.

In fluorescence video-microscopy, methods that estimate trajectories of small objects of interest (chromosomes, vesicles, ...) may encounter difficulties if the number of objects is large and the signal-to-noise ratio is low. Moreover, the tracked objects are not always visible in the sequence when tagging molecules separate suddenly from the target objects. Obviously, the complexity of dynamical processes involving many objects or groups of objects in interaction cannot be easily modeled. The corpus of data to be considered for a comparative analysis in a single experiment formed by multiple image series, is also massive. Nevertheless, it is now clear that the localization and spatio-temporal conformation of a large number of molecular constructions within the cell, their dynamical response to diverse chemical, physical or bio-molecular perturbations, are key elements for understanding the essential functional mechanisms in life sciences. Motion information and trajectories have to be extracted in order to analyze the dynamical response of the cell to different perturbations and experimental conditions.

In this paper, we propose a simulation/estimation framework able to model complex data corresponding to interactions between moving particles with variable velocities. Parsimonious models representing fluorescence microscopy image sequences will be defined to summarize complex data into a low dimensional set of parameters. These models will be exploited to generate artificial image sequences that mimic dynamics observed in real image sequences. In our study, the acquisition rate of a real image sequence is typically one stack per second at most. The volume/stack being is composed of 10 slices of 512×512 pixels. This constitutes standard settings in fluorescence wide-field microscopy. The speed of the vesicles ranges from 1 to 10 pixels and the number of objects can be large (about a few hundreds).

Traditionally, tracking algorithms compute object trajectories that have to be analyzed. Unlike previous methods (Smal et al., 2007; Genovesio et al., 2006), we need to simultaneously estimate the traffic component (e.g. moving objects) and the cytosolic component, both involved in membrane transport. The main difficulty is that these two adding components are perturbed by noise and photo-bleaching. Our goal is then to robustly estimate each factor and component for analysis. We introduce the simulation framework to evaluate the performance of estimation methods we propose.

1.2 Needs for simulation tools

In many application fields such as medical imaging or astronomy, simulations are required for validating physical models and understanding recorded data. In this section, we explain the rationale for simulation methods in video-microscopy.

First, realistic simulations of dynamical processes usually give a qualitative and controlled representation of the observed spatio-temporal biological events. Simulation can be then considered as a computational tool that can help to understand some mechanisms of internal components within the cell. By interacting with the control parameters, an expert can artificially simulate processes close to the reality provided the dynamical models are known ; this philosophy has been successfully exploited to understand dynamics of microtubule networks (Gibbons et al., 2001; Nédélec, 2001). By minimizing the difference between a set of descriptors computed from a real image sequence and the same set of descriptors computed from a simulated sequence, the parameters of the simulation method can be tuned to obtain an artificial sequence that reveals apparently the same dynamical characteristics than the observed sequence. This set of estimated control parameters can then be considered a parsimonious representation of the underlying process.

Moreover, dynamical information extraction usually relies on tasks such as object detection, motion estimation or object tracking. The most commonly used tracking concept is the so-called "connexionist" approach (Anderson et al., 1992; Sbalzarini and Koumoutsakos, 2005; Bonneau et al., 2005; Racine et al., 2006) which consists in detecting particles independently in each frame in a first step, and then linking the detected objects over time. The related *data association* task is the most critical step in this approach, especially if the number of objects is very high and if the trajectories interact. Sophisticated particle filtering techniques (Smal et al., 2007; Genovesio et al., 2006; Li et al., 2007) or graph-theory based methods (Thomann et al., 2003) have been then developed to improve temporal matching. These tasks cannot be done manually, and they must be fast, reliable and reproducible. Furthermore, comparing

object tracking results to ground truth is the more straightforward method to assess the performance of the applied method. Accordingly, simulation of a reliable ground truth is an important and challenging task especially in biomedical imaging. Let us point out that benchmarking data sets are for instance widely used to compare methods in image restoration (Portilla et al., 2003) and optical flow estimation (Barron et al., 1994). In fluorescence video-microscopy, the proposed simulation methods used to build benchmarking data sets are limited yet since they are not able to represent complex interactions between objects as observed in real image sequences. Nevertheless, in (Hadjidemetriou et al., 2006), the authors proposed to estimate the dynamics of outer tips of microtubules ; the method is validated on artificial data that mimic real image sequences. As for vesicle tracking within living cells, *random walks* combined with parametric models are commonly used for validation (Genovesio et al., 2006), but they cannot account for the complex movements of real moving objects in video-microscopy sequences.

1.3 Simulation framework and properties

Let us briefly discuss the expected properties of a tool to perform simulation of image sequences:

(1) Two modeling approaches can be proposed for simulation: *data-driven modeling* and *physically-based modeling*. The physics-based approach relies on the physical properties of the scene and the optical characteristics of the imaging system for image modeling. The main advantage is that the model parameters are motivated by physics. Hence, they are easy to interpret because they directly correspond to the real world. Conversely, the complexity of scenes and models usually limits such an approach and the *inverse problem* cannot be easily solved. The *data-driven modeling* aims at describing image sequences through statistical models learned from real images (Soatto et al., 2001). This approach can only mimic dynamical processes but is not able to describe the physical properties of real processes. *Data-driven* and *physically-based* approaches can be also combined to model the main components of the image sequence. In video-microscopy, these components are essentially the moving objects, the fixed or slowly-varying background and noise.

(2) A simulation method must also be *controllable* (Wang and Zhu, 2003). This means that the representation must be parsimonious, which is useful for interpretation by an expert. In most cases, the parameters are related to the physical properties of the system but also to the properties of the object image like scale or velocity. By using such a representation, the simulation method becomes more interactive and allows the expert to exploit *a priori* knowledge or to plan a set of experiments by editing the

simulation. For example, an expert can indicate the locations of source and destination points of moving objects, and by varying the positions of these extremity points, she/he can observe the evolution of the simulated intracellular trafficking. Finally, the expert feedback can be used to set up a realistic simulation.

Our aim is to fulfill these requirements.

1.4 Our approach

In this paper, we propose a powerful method for simulating complex video-microscopy image sequences. We design a realistic image sequence modeling framework able to mimic the dynamical and photometric contents of video-microscopy image sequences showing trafficking. Unlike the biophysical approach which aims at describing the underlying physical phenomena (Gibbons et al., 2001; Nédélec, 2001), the proposed approach is only based on the analysis of original image sequences. While being quite general, the proposed method has been designed for analyzing the role of fluorescence-tagged proteins moving around the Golgi apparatus and participating in the intracellular trafficking. These proteins can be linked to vesicles. The vesicles are propelled by motor proteins moving along polarized "cables" called microtubules, that form a dense network. This mechanism explains the observed high velocities which could not be accounted by basic diffusions. In order to model the content of these sequences, it is decomposed in two components. The first one is the almost static background of the scene while the second one contains the objects of interest, that is the tagged vesicles moving with high velocities. This representation yields a compact description of the dynamical processes corresponding to small moving objects within the cell.

The remainder of this paper is organized as follows. In Section 2, a dynamical background model is proposed and a method is defined to estimate the model parameters. In Section 3, a photometric and network-based dynamical model is introduced to represent moving spots in fluorescence microscopy image sequences. A statistical method is also presented for moving spot detection. Finally, in Section 4, we report several experimental results and demonstrate the potential of the proposed approach.

2 Dynamical background modeling

In this section, we propose a statistical framework for modeling and estimating the time-varying background.

2.1 Image model

Large structures lying inside the cell like the Golgi apparatus appear as nearly static during the observation time interval. In images showing fluorescently tagged particles, the global image intensity is proved to vary slowly along time. This is due to several physical phenomena such as photo-bleaching or diffusion of fluorescent proteins within the cell. Therefore, it is appropriate to propose a model able to describe the (slowly) time-varying background since a stationary model would be too restrictive. The modeling of more complex small moving objects with variable velocities (traffic component) will be discussed in Section 4.

We have conducted experiments showing that the intensity variation with respect to time can be well captured by a linear model for each pixel of the image grid, mainly since we are dealing with sequences of limited length. This simple model provides a compact representation of the background. For each point, it amounts to two parameters (linear temporal model). Consequently, the background can be fully described by two 2D maps corresponding to the two spatially varying parameters. The processing of 3D volume sequences yields 3D maps if considered. Nevertheless, the involved parameters are spatially correlated, which must be taken into account in the estimation process. Let us point out that the proposed approach, explained below, can be adapted to non-linear intensity models (exponential or bi-exponential models) in 2D or 3D, if desired.

Formally, we propose the following image sequence model for the background:

$$f(\mathbf{x}, t) = a(\mathbf{x}) + b(\mathbf{x})t + u(\mathbf{x}, t) + \epsilon(\mathbf{x}, t) \tag{1}$$

where $f(\mathbf{x}, t)$ denotes the intensity observed at pixel $\mathbf{x} = (x, y)^T \in \Omega$ (or $\mathbf{x} = (x, y, z)^T$ for 3D volume sequences) and time t where Ω denotes the 2D or 3D image support. The two coefficients $a(\mathbf{x})$ and $b(\mathbf{x})$ varies with the spatial image position and $u(\mathbf{x}, t)$ is a positive function that describes the intensity of moving vesicles if any. In the sequel, $\epsilon(\mathbf{x}, t)$ is an additive white Gaussian noise. This model is able to describe the background intensity of the whole image sequence with only two 2D/3D maps $\mathbf{a} = \{a(\mathbf{x})\}$ and $\mathbf{b} = \{b(\mathbf{x})\}$ of the same size as the image grid. In the next section, we describe a method to estimate the maps $\mathbf{a}$ and $\mathbf{b}$ representing the time-varying background model.

2.2 Pixel-wise estimation of the background model parameters

We first deal with the estimation of parameters $a(\mathbf{x})$ and $b(\mathbf{x})$ at a given location $\mathbf{x}$, *i.e.*, for a single temporal 1D signal. Let us stress out that this

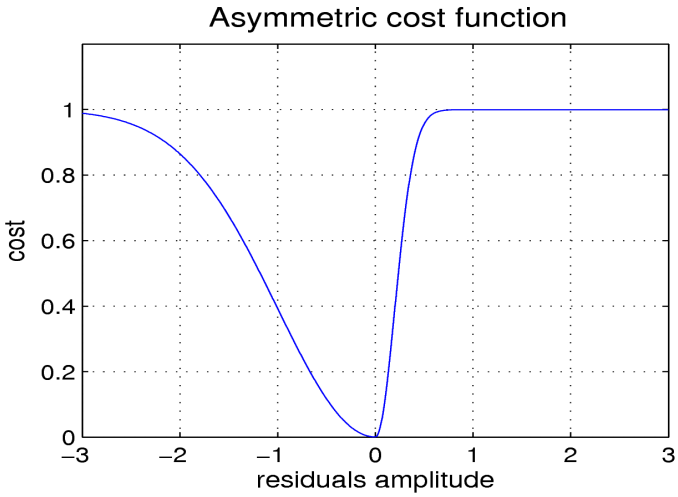

Fig. 1. The Asymmetric Leclerc robust function allows us to better discard outliers related to moving vesicles when estimating the background component.

point-wise estimation must be performed several millions of times (for each image point) if the method is applied to a 3D image sequence. Accordingly, the proposed estimation procedure must be very fast. Besides, in our study, vesicles have an erratic behavior and sometimes stop for a long time. Consequently, prior motion detection cannot be used to extract these objects from the background. The estimation of the time-varying background will be then based on the image intensity only. Also, since the background estimation can be altered by the presence of moving vesicles, we will resort to a robust estimation framework.

2.2.1 Robust M-estimation

The two parameters $a(\mathbf{x})$ and $b(\mathbf{x})$ are estimated by minimizing a robust error function of the form

$$\mathcal{E}(a(\mathbf{x}), b(\mathbf{x})) = \sum_{t=1}^{n} \rho\left(f(\mathbf{x}, t) - (a(\mathbf{x}) + b(\mathbf{x})t)\right), \tag{2}$$

where n is the number of temporal samples in the 1D signal and $\rho(\cdot)$ is a robust function. A local minimum of $\mathcal{E}(a(\mathbf{x}), b(\mathbf{x}))$ is commonly obtained by using the iteratively re-weighted least squares (IRLS) procedure (Huber, 1981).

The choice of the robust function $\rho(\cdot)$ is usually guided by the noise probability density function (Ieng et al., 2004). In our case, the overall noise is the sum of two components $u(\mathbf{x}, t)$ and $\epsilon(\mathbf{x}, t)$. In order to take into account that $u(\mathbf{x}, t)$ usually takes high positive values (vesicles appear as bright spots in the image), we choose an asymmetric robust function, more specifically, the Leclerc estimator (Allende et al., 2006; Ruckstuhl et al., 2001) plotted in

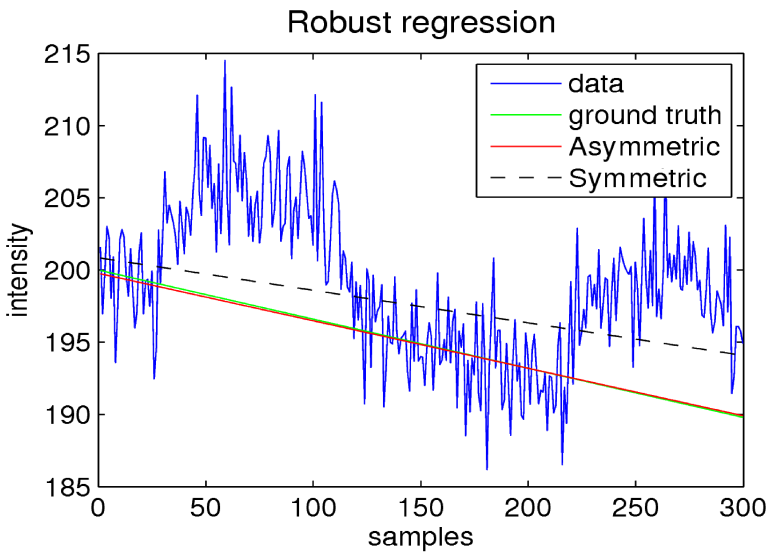

Fig. 2. Regression using the asymmetric robust Leclerc function and the symmetric one. The asymmetric estimator (red) fits well the ground truth (green), while the symmetric function provides biased results (dotted line).

Figure 1 and defined as:

$$\rho(z) = \begin{cases} 1 - \exp\left(-\dfrac{z^2}{\lambda^2\sigma_1^2}\right) & \text{if} \quad z \leq 0, \\ 1 - \exp\left(-\dfrac{z^2}{\lambda^2\sigma_2^2}\right) & \text{otherwise.} \end{cases} \tag{3}$$

The scale σ_2 factor can be estimated by applying a robust least-trimmed squares (LTS) estimator to the *pseudo-residuals* (Gasser et al., 1986) defined as: $s(\mathbf{x}, t) = (f(\mathbf{x}, t+1) - f(\mathbf{x}, t))/\sqrt{2}$, where the coefficient $1/\sqrt{2}$ ensures that $\mathbb{E}[(s(\mathbf{x}, t))^2] = \mathbb{E}[(f(\mathbf{x}, (t))^2]$. The scale factor σ_1 is estimated by using the variance of the residuals given by the least-mean squares estimator and obtained at the initialization step. Let us point out that, in regions where there are no moving vesicles, σ_1 and σ_2 are found almost equal. The scale parameter λ acts as a threshold and is chosen in the range $[1, 3]$. Theoretically, this value is relative to the point where the derivative of the $\rho'(\cdot)$-function is zero (Black et al., 1998).

As a matter of fact, the proposed estimator is biased (Ruckstuhl et al., 2001) but the bias is small. Simulations proved that the L_2 risk of the estimator is smaller when an asymmetric cost function is used and when the data are corrupted by an additive positive signal. Figure 2 shows that the proposed estimator is able to deal with heavily contaminated data and outperforms the symmetric Leclerc M-estimator.

2.2.2 Confidence matrix

An accurate estimation of the confidence matrix for the estimated parameters is needed for the subsequent steps described in Section 2.3. We use the approximation proposed in (Ieng et al., 2004) to compute the covariance matrix of the estimator

$$\widehat{C}(\mathbf{x}) = \frac{\sum_{t=1}^{n} w(r(\mathbf{x},t))(r(\mathbf{x},t))^2 \sum_{t=1}^{n} (w(r(\mathbf{x},t)))^2}{\left(\sum_{t=1}^{n} w(r(\mathbf{x},t))\right)^2} \times \begin{pmatrix} \sum_{t=1}^{n} w(r(\mathbf{x},t)) & \sum_{t=1}^{n} w(r(\mathbf{x},t))t \\ \sum_{t=1}^{n} w(r(\mathbf{x},t))t & \sum_{t=1}^{n} w(r(\mathbf{x},t))t^2 \end{pmatrix}^{-1} \tag{4}$$

where $r(\mathbf{x},t) = f(\mathbf{x},t) - (a(\mathbf{x}) + b(\mathbf{x})t)$ and the weights are defined as $w(z) = \rho'(z)/z$. Unlike the expression given in (Huber, 1981), the approximation given by (4) is not asymptotic and yields a better estimation of the covariance matrix when n is small (e.g., $n < 500$).

2.3 Spatial coherence for background estimation

In this section, we introduce a process to regularize the 2D/3D maps $\mathbf{a}$ and $\mathbf{b}$. This can be accomplished by adopting the *bias-variance trade-off* framework described in (Lepski, 1991; Maurizot et al., 1995; Ercole et al., 2005; Kervrann and Boulanger, 2006). Instead of using a single temporal signal at each location $\mathbf{x}$ to estimate $a(\mathbf{x})$ and $b(\mathbf{x})$, a set of temporal 1D signals is first collected in a spatial neighborhood of pixel $\mathbf{x}$. This collection of signals is then analyzed in order to take into account the desired spatial coherence of the parameters. In practice, a finite set of nested space-time tubes/parallelepipeds is considered whose temporal section is formed by a growing spatial neighborhood centered at point $\mathbf{x}$ (see Figure 3). Each tube $\mathcal{T}_l(\mathbf{x})$ centered at point $\mathbf{x}$ is parameterized by its diameter $\phi_l(\mathbf{x})$ and $l \in [1, .., L]$ denotes the index associated to each tube defined as

$$\mathcal{T}_l(\mathbf{x}) = \{\mathbf{y} \in \Omega : \|\mathbf{x} - \mathbf{y}\|_p < \phi_l(\mathbf{x})\}, \tag{5}$$

where $\|\cdot\|_p$ denotes the L_p norm. In our experiments, we arbitrarily choose the L_∞ norm to design the set of nested parallelepipeds shown in Fig. 3.

In order to select the optimal diameter of the space-time tube, we propose to minimize the point-wise L_2 risk $\mathbb{E}[(\widehat{\theta}(\mathbf{x}) - \theta(\mathbf{x}))^2]$ of the parametric estimator where $\theta(\mathbf{x}) = (a(\mathbf{x}), b(\mathbf{x}))^T$ is the true parameter vector and $\widehat{\theta}(\mathbf{x})$ its corresponding estimator, at position $\mathbf{x}$. The L_2 risk can be decomposed into the squared bias and the variance. As shown in Figure 4, while the diameter $\phi_l(\mathbf{x})$ increases with l, the bias increases too. This can be explained by the fact that the data cannot be described any longer at same stage by a unique parametric model. In contrast, by taking more data points, the variance decreases. This

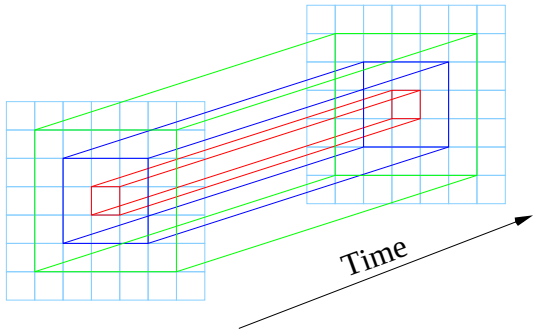

Fig. 3. Set of nested tubes $\{\mathcal{T}_{i,l}\}_{l=1,\cdots,3}$. Respectively red, blue, green parallelepipeds are successively considered if we choose the L_∞ norm (see text).

behavior, also called *bias-variance trade-off*, is exploited to detect the minimum of the point-wise L_2 risk which is nearly equal to twice the variance (Lepski, 1991) (see Fig. 4).

For each diameter $\phi_l(\mathbf{x})$, new estimates of the background model parameters $\widehat{\theta}_l(\mathbf{x})$ and the associated covariance matrix $\widehat{C}_l(\mathbf{x})$ are computed with the procedure described in Section 2.2 but using now all the data contained in the considered neighborhood. It can be shown that the *bias-variance trade-off* writes down as the following test (Kervrann and Boulanger, 2006):

$$\frac{n-2+1}{2n}\left(\widehat{\theta}_l(\mathbf{x})-\widehat{\theta}_{l'}(\mathbf{x})\right)^T \widehat{C}_{l'}^{-1}(\mathbf{x})\left(\widehat{\theta}_l(\mathbf{x})-\widehat{\theta}_{l'}(\mathbf{x})\right) < \eta \tag{6}$$

for all $1 \leq l' < l$. While this inequality is satisfied, the diameter of the tube is increased and the estimation process is continued. It is established that the threshold η can be defined as a quantile of a Fisher distribution of parameters 2 and $n-2-1$, since an estimator of covariance matrix is only available.

In this section, we have proposed a spatially and temporally varying background model and a statistical framework to estimate the involved model parameters. In the second part of the paper, we develop a simulation framework to generate dynamical content corresponding to small moving spots in image sequences.

3 Spot model

In video-microscopy, vesicles appear in many image sequences as small bright spots against a dark background. The object diameter theoretically ranges from 60 nm to 150 nm. The resolution of the microscope is about $130\times130\times300$

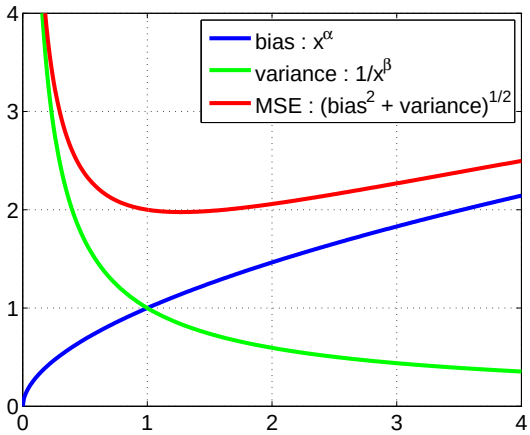

Fig. 4. *Bias-variance trade-off* principle. When the tube diameter increases, the bias increases and the variance decreases. The optimum is achieved when the bias and the variance are of the same order.

nm. Then, the diameters of spots are often below this spatial resolution. However, the point spread function of the video-microscope makes them appear as larger structures even if a deconvolution process is applied (Sibarita et al., 2002). Furthermore, when the density of objects increases, vesicles gather together and constitute small rods.

These vesicles are also known to move along microtubules, that is along polymers that have an exceptional bending stiffness and can be easily fit by smooth curves. Microtubules are conveyor belts inside the cell. They drive vesicles, granules, organelles like mitochondria, and chromosomes with special attachment proteins using molecular motors. It is also established that molecular motors form a class of proteins responsible for the intracellular transport inside the cell (White et al., 1999). The dynein and kinein proteins are two classes of motors associated with microtubules. It has been shown that the concentration of these molecular motors influences the structure and the dynamics of the microtubule network (Nédélec, 2001). In stable conditions, the speed and polarization of these motors is assumed to be constant. This explains partially why the observed velocity of vesicles is constant if they move along the same microtubule.

In our study, vesicles move along the microtubule, leaving a donor organelle and reaching an acceptor organelle, e.g., from the Golgi apparatus to the Endoplasmic Reticulum. Unlike (Hadjidemetriou et al., 2006), we assume that the microtubule network is static when compared to moving vesicles and we rather aim at generating video-microscopy image sequences following a *data-driven* approach.

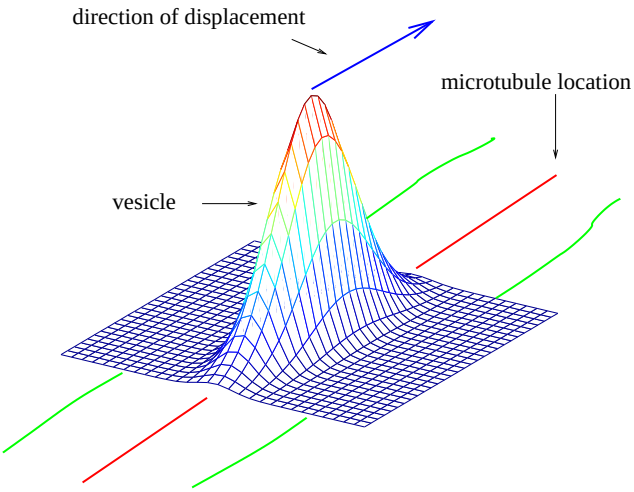

Fig. 5. Gaussian model of the spot oriented in the direction of the microtubule axis. The covariance matrix of the Gaussian function depends on the velocity of the vesicle. The simulated vesicles are then elongated along with the displacement direction.

3.1 Photometric model

Large vesicles or sets of nearby vesicules can be satisfyingly represented by anisotropic Gaussian spots with variances related to the spot dimensions ranging from 60 nm to 150 nm. Furthermore, the size of vesicles in the image is close to the pixel size. The size of spots will be estimated in the image sequence as described in Section 3.3. In our approach, occlusions are not handled since tagged molecules attached to vesicles can diffuse over time. In addition, split and merge processes involving several vesicles occur in real image sequences leading to rods (Zhang et al., 2006). The latter can be considered as a unique object since the merged spots move along the same microtubule.

In what follows, the covariance matrix of the anisotropic Gaussian spot is a function of the displacement direction. The ellipticity also depends on the velocity. Figure 5 displays how the covariance matrix of the anisotropic Gaussian function allows us to modify the orientation of spots according to the direction of the microtubule axis.

3.2 Dynamical model for traffic

3.2.1 Microtubule network modeling

A physics-based simulation of the self-organization of the microtubule network can be found in (Surrey et al., 2001). It is based on the interaction between the motors (e.g., kinesine) and microtubules, and explains some characteristic conformations such as mitotic spindle. Typically, it takes into account the dynamical behavior of the microtubules. However, this computer simulation only describes the behavior of the microtubule network *in-vitro* and is not adapted for the more complex *in-vivo* case in which the microtubules interact with other organelles of the cell. In addition, the observation time intervals are usually short compared to the dynamics of the network itself, assumed to be fixed in the following.

In order to generate a synthetic but realistic microtubule network, we exploit real image sequences as input for the modeling. A real network could be tagged with Green Fluorescence Protein (GFP) but this network is too complex and individual microtubules cannot be easily extracted. However, the microtubule network can be also coarsely computed from a maximum intensity projection map with respect to time, that is from the paths followed by the tagged vesicles. For instance, Fig. 6 shows the maximum intensity projection map of a real sequence of 300 2D images. This simple projection allows us to select a subset of the main paths used for the intracellular trafficking, leading to a network with low complexity. This approach has been successfully exploited for the construction of kymograms in (Sibarita et al., 2006). However, as shown in Fig. 6, all the paths are not complete, especially if the sequence duration is too short. The gaps are then completed by using *ad-hoc* image processing tools. The locations of the roads are extracted from the network image using the *unbiased line detection* algorithm defined in (Steger, 1998). Finally, each road is finally described by its length, its width, its source node and its destination node.

3.2.2 Selection of source/destination nodes

In the proposed simulation, vesicles are going from one region to another. Typically, they leave a donor organelle and move toward an acceptor organelle. Once the network has been computed, the expert needs to specify the source and destination nodes on the network. In order to take into account the lack of *a priori* information on the organelles and their function, a node can be both a source and a destination, while the other nodes represent the intersection points of the network and are only used for routing. Source-destination pairs are important cues for the simulation and corresponds to a birth/death map

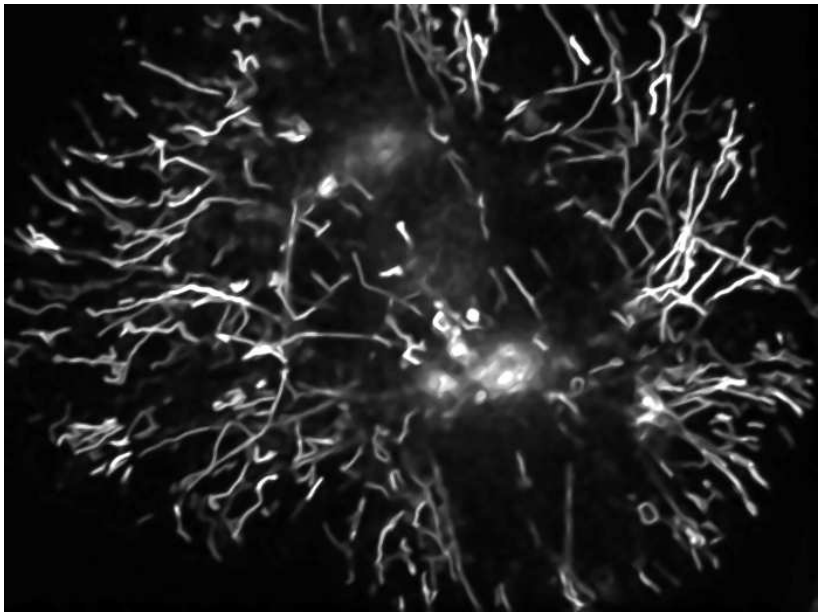

Fig. 6. Maximum intensity projection map computed from an image sequence. The paths followed by the vesicles appear as bright filaments. The maximum intensity projection map has been simplified using the denoising method described in (Kervrann and Boulanger, 2006).

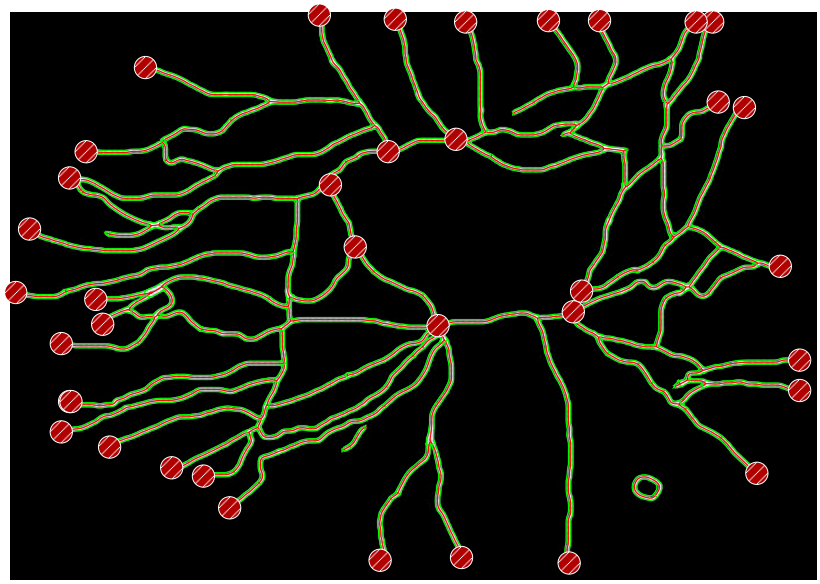

Fig. 7. Representation of a realistic synthetic network. This network is based on a maximum intensity projection map and has been manually simplified. This network is composed of 146 nodes and 160 bi-directional links which correspond to 320 directional edges in the graph associated to the network.

as described in (Wang and Zhu, 2003). These labels are actually related to the locations and relationships of specific organelles inside the cell.

In Fig. 7, the source and destination nodes have been manually selected by the expert. The destination nodes displayed in red correspond to "end-points", while the source nodes corresponding to the membrane of the Golgi apparatus are displayed in green. In this simulation, vesicles are only going from the Golgi to the "end-points" located at the periphery of the cell. Thus, the retrograde transport from "end-points" to Golgi is prohibited and assumed to be inhibited by bio-chemical alterations.

In our approach, the paths are defined as the minimal paths between the source and the destination nodes. They are computed using the Dijkstra algorithm (Dijkstra, 1959). In that case, the weight associated to each edge can be defined as a function of the length of the corresponding road, but it could take into account other parameters as well. Let us note for instance the allowed speed for a given edges could be retained to estimate the shortest path. Finally, as expected, the vesicles move along the estimated roads with velocities distributed around the speed-limit of the roads. At each time step, the vesicle is moved along the microtubule with a displacement step which is a proportional to the velocity.

3.3 Estimation of model parameters and spot detection

Given a simulated sequence as described above, we now address the problem of detecting moving vesicles with minimal prior knowledge. The performance of any detector is better assessed if ground truth is available, which is precisely what this simulation framework can offer. Blob/spot detection in image sequences is an important task in video-microscopy. Here, we propose a method able to decide with high confidence whether any image point belongs to the image background or a moving object. Our approach is unsupervised. It will be applied to artificial image sequences obtained by the modeling framework described in the paper and to real image sequences.

Our two-step method for temporal detection is based on the minimization of a penalized likelihood criterion in the line of work of Birgé and Massart (Birgé and Massart, 2001) for model selection. In this step, we consider a sequence of n observations $Y_1, \cdots, Y_n$ (assumed to be i.i.d.) that take its values in $\mathbb{R}^+$. We assume that a subset of $n-\ell, \ell \in \{0, \cdots, n\}$, observations are realizations of a hidden variable m at some unknown instants. Here K_ℓ will denote the unknown number $n-\ell$ of temporal observations defined at time t as:

$$Y_t = m + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \tau^2). \tag{7}$$

where m and the variance τ^2 are also unknown. To detect the K_ℓ observations corresponding to this model, we adopt a global approach where all the relevant observations are simultaneously detected, by minimizing a penalized

least squares criterion. The role of the least squares term is to measure the fit of of some observations Y_t's to m the more accurately possible. The penalty term is used for determining the number of observations used to estimate m and τ^2.

In our application, a subset of residuals $Y_t \triangleq f(\mathbf{x}, t) - (a(\mathbf{x}) + b(\mathbf{x})t)$ (obtained as explained in Section 2.2.1) is assumed to belong to the dark background of intensity m at pixel $\mathbf{x}$. The remaining observations are assumed to be related to the moving objects. For the sake of simplicity in the notations, we will omit the variable $\mathbf{x}$ since the detection is performed independently at each pixel. Without loss of generality, the sequence of residuals $Y = \{Y_1, \cdots, Y_n\}$ is first re-ordered to produce a new sequence $Y' = \{Y'_1, \cdots, Y'_n\}$ such that $|Y'_t| > |Y'_{t+1}|$ since the Y_t's are assumed to be independent. Now, we consider the following collection of models $\mathcal{M} = \{M_1, ..., M_n\}$:

$$\begin{cases} M_1 & = (Y'_1,\, 0,\quad 0,\quad ...,\, 0)^T \\ & \vdots \\ M_{n-1} & = (Y'_1,\, Y'_2,\, Y'_3,\, ...,\, 0)^T \\ M_n & = (Y'_1,\, Y'_2,\, Y'_3,\, ...,\, Y'_n)^T. \end{cases} \tag{8}$$

where M_ℓ is a n-dimensional vector and $K_\ell = n - \ell$ is the number of components of the background . Since the number of components of the background is unknown, we introduce a penalized version of the least squares criterion for signal decomposition (Huet, 2006), defined as :

$$\mathcal{J}(M_\ell) = \frac{n}{2}\log(\widehat{\tau}^2_\ell) + n\left(c_1 \log\left(\frac{n}{n-K_\ell}\right) + c_2\right)\frac{n-K_\ell}{K_\ell} \tag{9}$$

where

$$\widehat{\tau}^2_\ell \;=\; \frac{\|Y' - M_\ell\|^2}{n} \;=\; \frac{1}{n}\sum_{t=\ell}^{n} Y_t'^2, \tag{10}$$

Y' is the set of n re-ordered observations . The two universal constants c_1 and c_2 were calibrated in (Huet, 2006) and found to be $c_1 = 2$ and $c_2 = 4$. Since the number n of observations is small, we compute $\mathcal{J}(M_\ell)$ for each model M_ℓ and select the model $M_\star$ such that:

$$M_\star = \inf_{M_\ell, l=1,\cdots,n} \mathcal{J}(M_\ell), \tag{11}$$

for every pixel in the image. This procedure amounts to thresholding the sequence of residuals (Huet, 2006).

In the second step, we collect the estimated models $M_\star(\mathbf{x})$ at each pixel $\mathbf{x} \in \Omega$. Since the model selection amounts to determining a threshold $T(\mathbf{x})$ such that $T(\mathbf{x}) = Y_{n-K_\star}(\mathbf{x})$ for the 1D sequence of residuals, a threshold map T is

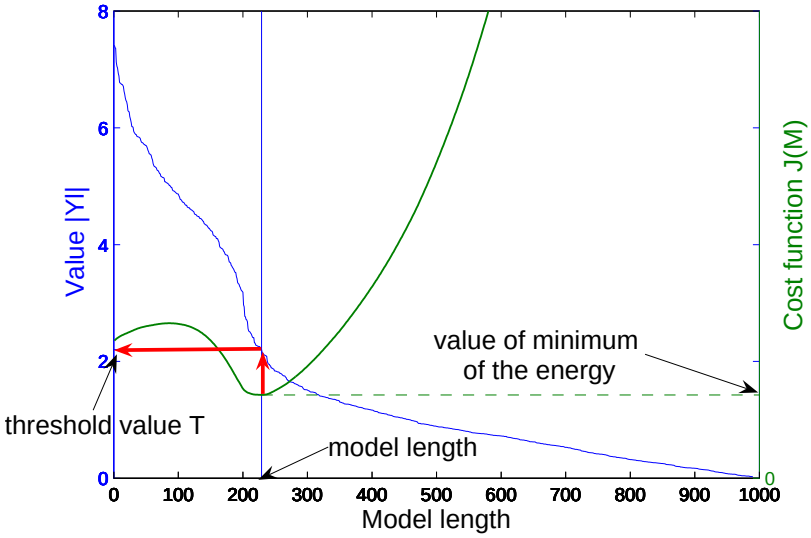

Fig. 8. Illustration of the *model selection* principle applied to a synthetic signal ($n = 1000$). The number of samples corresponding to the background is about 800. The X-axis represents the model number ; the left ordinate corresponds to the value $|Y_\ell|$ associated to the model M_ℓ plotted in blue ; the right ordinate represents the value of the cost function $\mathcal{J}(M_\ell)$.

computed. Finally, we determine a global threshold γ for the whole image in order to detect the highest intensity level corresponding to the background component for the whole image sequence (see Fig. 8). Practically, we use the model selection procedure described earlier to threshold the T map. In that case, the threshold values $T(\mathbf{x})$ are considered as i.i.d. observed variables. Finally, a pixel $\mathbf{x}$ belongs to the background if $f(\mathbf{x}, t) - (a(\mathbf{x}) + b(\mathbf{x})t) \leq \gamma$ and to moving objects otherwise.

4 Experiments

In this section, we report two experiments to illustrate the proposed simulation method. First, we show how we can generate a realistic image sequence from real images. Second, a synthetic image sequence whose parameters have been manually determined is exploited to evaluate the performance of an image denoising method as well as the estimation method of the background as described in Section 3.

4.1 Realistic image sequence simulation

We utilize a real image sequence in order to simulate a sequence with similar and controlled photometric and dynamical content. One frame of the real image is shown in Fig. 9(a) and contains vesicles moving from the Golgi apparatus to the Endoplasmic Reticulum. The parameters of the time-varying background are first estimated as described in Section 2. The two 2D maps $\mathbf{a}$ and $\mathbf{b}$ are respectively shown in Fig. 9(e) and in Fig. 9(f). Once these parameter maps have been estimated, the background is subtracted from the original images in order to obtain the sequence of residuals displayed in Fig. 9(g), which is a noisy representation of the moving spots. The main paths followed by vesicles through the 150 frames of the real sequence can be observed on the maximum intensity projection map in the time direction as shown in Fig. 9(b). We have further enhanced the maximum intensity projection map using optimal steerable filters (Jacob and Unser, 2004) (Fig. 9(c)). The *unbiased line detection* algorithm (Steger, 1998) is then applied to the resulting enhanced image in order to estimate the positions of the roads plotted in Fig. 9(d). Finally 150 vesicles are generated and moved along the reconstructed network. The velocities of the vesicles are tuned so that the simulated sequence provides the same visual effect than the original sequence. It was confirmed by biologist-experts that the proposed simulation method supplies realistic image sequences both considering photometry and dynamics aspects.

4.2 Benchmark generation

This simulation framework can also be exploited for benchmarking. More specifically, in order to objectively evaluate the performance of the proposed background estimation method, we have simulated a $128 \times 128 \times 150$ image sequence. This simulation comprises 20 vesicles moving along the network plotted in Fig. 10(a). To generate the background model, we have manually designed the spatial appearance of the background as shown in Fig. 10(b). Then, we have computed the image such that after a period equals to twice the duration of the simulation, we would obtain a uniform flat background. The two artificial maps $\mathbf{a}_0$ and $\mathbf{b}_0$ for the background model are then shown in Fig. 10(b)-(c). Three frames of the noise-free simulated sequence are given in Fig. 10(d)-(f). A zero-mean Gaussian noise of standard deviation $\sigma = 9$ has been added to these frames and the resulting noised images are displayed in Fig. 10(i,j,k). The intensity of the vesicles are assumed to follow a Gaussian law with mean 30 and standard deviation 3 leading to a signal-to-noise ratio of 13.6dB. By combining all the mentioned inputs, we have specified the useful set of parameters to control the simulation.

Now, we apply the algorithm described in Boulanger et al. (2008, 2007) to restore the noisy image sequence obtained as explained above. The two-step denoising method exploits 3D+time information to improve the signal-to-noise ratio of images corrupted by mixed Poisson-Gaussian noise (Boulanger et al., 2008, 2007). A data-driven variance stabilization transform is first applied to the image-data to introduce independence between the mean and variance. In a second step, an original statistical patch-based framework for noise reduction and preservation of space-time discontinuities has been developped. In our study, discontinuities are related to small moving spots with high velocity observed in fluorescence video-microscopy. The idea is to minimize an objective nonlocal energy functional involving image spatio-temporal patches. The minimizer has a simple form and is defined as the weighted average of input data taken in spatially-varying neighborhoods. The size of each neighborhood is optimized to improve the performance of the pointwise estimator. By running 6 iterations of the adaptive denoising algorithm (3×3 patches), the resulting mean squared error of the recovered image is 1.35, which corresponds to a signal-to-noise ratio of 30.0dB. The noise is therefore drastically reduced (see Fig. 11) and, visually the reconstructed image sequence looks similar to the original sequence. Let us also stress that no vesicle has disappeared. Denoising algorithms are commonly applied to real image sequences in microscopy to make easier other image processing tasks.

We have evaluated the performance of our estimation method of the time-varying background model. This method is applied to a simulated noisy image sequence. The two estimated maps $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$ are shown in Fig. 11. The mean squared error between the original map $\mathbf{a}_0$ and the estimated map $\widehat{\mathbf{a}}$ is 1.50 and the signal-to-noise ratio is 31.44dB. The signal-to-noise ratio calculated from the original map $\mathbf{b}_0$ and the estimated map $\widehat{\mathbf{b}}$ is 32.70dB. The results of the temporal detection method described in Section 3.3, are presented in Fig.11(j,k,l) at time $t = 0, t = 75$ and $t = 100$.

5 Conclusion

In this paper, we have proposed a framework for the analysis and the simulation of the dynamical content corresponding to membrane trafficking in fluorescence video-microscopy. We have designed models for the time-varying background and moving vesicles. We have also proposed statistical methods for estimating the model parameters. The proposed simulation framework has been demonstrated on artificial and data-driven image sequences.

More generally, the simulation framework can be used to generate realistic image sequences in fluorescence time-lapse microscopy. Any tracking, object detection and segmentation algorithm can be evaluated since the ground truth

is available. Since it is challenging to assess the performance of image processing algorithms on real image sequences, our motivation was to determine a priori the limits of each algorithm (based on metrics to be arbitrarily defined such as mean square error, false alarm rate, ...). When applied to real images, the results can be better quantified if well known on artificial images.

Nevertheless, further validations with biologist-experts are required to improve the proposed modeling framework. As already investigated in(Pécot et al., 2007, 2008), we also plan to address the traffic estimation problem based on the image sequence modeling we have described in this paper.

References

Allende, H., Frery, A., Galbiati, J., Pizarro, L., Nov. 2006. M-estimator with asymmetric influence function: the $\mathcal{G}_a^0$ distribustion case. Journal of Statistical Computation and Simulation 76 (11), 941–946.

Anderson, C., Georgiou, G., Morrison, I., Stevenson, G., Cherry, R., 1992. Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charged-coupled device camera. low-density lipoprotein and influenza virus receptor mobility at 4 degrees c. Journal of Cell Science 101, 415–425.

Barron, J. L., Fleet, D. J., Beauchemin, S. S., Jan. 1994. Performance of optical flow techniques. International Journal of Computer Vision 12 (1), 43–77.

Birgé, L., Massart, P., 2001. Gaussian model selection. Journal of European Mathematical Society 3, 203–268.

Black, M., Sapiro, G., Marimont, D., Heeger, D., Mar. 1998. Robust anisotropic diffusion. IEEE Trans. on Image Processing 7 (3), 421–432.

Bonneau, S., Dahan, M., Cohen, L. D., Jun. 2005. Tracking single quantum dots in live cells with minimal paths. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'2005. Vol. 3. San Diego, CA, pp. 141–149.

Boulanger, J., Kervrann, C., Bouthemy, P., Jun. 2007. Space-time adaptation for patch-based image sequence restoration. IEEE Trans. on Pattern Analysis and Machine Intelligence 29 (6), 1096–1102.

Boulanger, J., Sibarita, J., Kervrann, C., Bouthemy, P., May 2008. on-parametric regression for patch-based fluoresence microscopy image sequence denoising. In: Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08). Paris, France.

Dijkstra, E. W., Dec. 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1 (1), 269–271.

Ercole, C., Foi, A., Katkovnik, V., Egiazarian, K., Jan. 2005. Spatio-temporal pointwise adaptive denoising in video: 3D non parametric approach. In:

Proc. of 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'2005. Scottsdale.

Gasser, T., Sroka, L., Jennen-Steinmetz, C., 1986. Residual variance and residual pattern in nonlinear regression. Biometrika 73, 625–633.

Genovesio, A., Liedl, T., Emiliani, V., Parak, W. J., Coppey-Moisan, M., Olivo-Marin, J.-C., May 2006. Multiple particle tracking in 3D+t microscopy: Method and application to the tracking of endocytosed quantum dots. IEEE Trans. on Image Processing 15 (5), 1062–1070.

Gibbons, F., Chauwin, J., Despósito, M., José, J., Jun. 2001. A dynamical model of kinesin-microtubule motility assays. Biophysical Journal 80 (6), 2515–2526.

Hadjidemetriou, S., Toomre, D., Duncan, J. S., 2006. Tracking the motion of the outer tips of microtubules. In: Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'2006). Washington, D.C., pp. 530–533.

Huber, P. J., 1981. Robust Statistics. Johm Wiley and Sons, New York, NY.

Huet, S., Mar. 2006. Model selection for estimating the non zero component of a gaussian vector. ESAIM P and S 10, 164–183.

Ieng, S.-S., Tarel, J.-P., Charbonnier, P., May 2004. Evaluation of robust fitting based detection. In: Proc. 8th Eur. Conf. on Computer Vision, ECCV'2004. Vol. 2. Prague, Czech Republic, pp. 341–352.

Jacob, M., Unser, M., Aug. 2004. Design of steerable filters for feature detection using canny-like criterion. IEEE Trans. on Pattern Analysis and Machine Intelligence 26 (8), 1007–1019.

Kervrann, C., Boulanger, J., 2006. Optimal spatial adaptation for patch-based image denoising. IEEE Trans. on Image Processing 15 (10), 2866–2878.

Lepski, O., 1991. Asymptotically minimax adaptive estimation 1: upper bounds. SIAM Journal Theory of Probability and Application 36 (4), 654–659.

Li, K., Chen, M., Kanade, T., Oct. 2007. Cell population tracking and lineage construction with spatiotemporal context. In: Proc. 10th International Conference on Medical Image Computing and Computer Assisted Intervention. Brisbane, Australia, pp. 295–302.

Maurizot, M., Bouthemy, P., Delyon, B., Iouditski, A., Odobez, J.-M., Oct. 1995. Determination of singular points in 2D deformable flow fields. In: Proc. IEEE Int. Conf. on Image Processing (ICIP'1995). Vol. 3. Washington, DC, pp. 488–491.

Nédélec, F., Sep. 2001. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. Journal of Cell Biology 158 (6), 1005–1015.

Portilla, J., Strela, V., Wainwright, M., Simoncelli, E., Nov. 2003. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. on Image Processing 12 (11), 1338–1351.

Pécot, T., Boulanger, J., Kervrann, C., Bouthemy, P., Apr. 2007. Network tomography for trafficking simulation and analysis in fluorescence microscopy imaging. In: Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to

macro (ISBI'07). Washington, D.C., pp. 268–271.
Pécot, T., Kervrann, C., Bouthemy, P., May 2008. Minimal paths and probabilistic models for origin-destination traffic estimation in live cell imaging. In: Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08). Paris, France.
Racine, V., Hertzog, A., Jouaneau, J., Salamero, J., Kervrann, C., Sibarita, J.-B., Apr. 2006. Multiple target tracking of 3D fluorescent objects based on simulated annealing. In: Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'2006). Washington, DC, pp. 1020 – 1023.
Ruckstuhl, A., Jacobson, M., Field, R., Dodd, J., Jan. 2001. Baseline subtraction using robust local regression estimation. Journal of Quantitative Spectroscopy and Radiative Transfer 68 (2), 179–193.
Sbalzarini, I., Koumoutsakos, P., 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. Journal of Structural Biology 151, 182–195.
Sibarita, J.-B., Magnin, H., De Mey, J. R., Jun. 2002. Ultra-fast 4D microscopy and high throughput distributed deconvolution. In: Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'2002). Washington, DC, pp. 769–772.
Sibarita, J.-B., Racine, V., Salamero, J., Apr. 2006. Quantification of membrane trafficking on a 3D cytoskeleton network in living cells. In: Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'2006). Washington, DC, pp. 263–266.
Smal, I., Draegestein, K., Galjart, N., Niessen, W., Meijering, E., 2007. Rao-blackwellized marginal particle filtering for multiple object tracking in molecular bioimaging. In: IPMI. pp. 110–121.
Soatto, S., Doretto, G., Wu, Y. N., Jul. 2001. Dynamic textures. In: Proc. 8th IEEE Int. Conf. on Computer Vision (ICCV'2001). Vol. 2. Vancouver, Canada, pp. 439–446.
Steger, C., Feb. 1998. An unbiased detector of curvilinear structures. IEEE Trans. on Pattern Analysis and Machine Intelligence 20 (2), 113–125.
Surrey, T., Nédélec, F., Leibler, S., Karenti, E., May 2001. Physical properties determining self-organization of motors and microtubules. Science 292, 1167–1171.
Thomann, D., Dorn, J., Sorger, P., Danuser, G., Sep. 2003. Automatic fluorescent tag localization ii: improvement in super-resolution by relative tracking. Journal of Microscopy 211 (3), 230–248.
Wang, Y., Zhu, S. C., Oct. 2003. Modeling textured motion: Particle, wave and sketch. In: Proc. 9th IEEE Int. Conf. on Computer Vision (ICCV'2003). pp. 213–220.
White, J., Johannes, L., Mallar, F., Girod, A., Stephan, G., Reinsh, S., Keller, P., Tzschaschel, B., Echard, A., Goud, B., Stelzer, E., Nov. 1999. Rab6 coordinates a novel golgi to ER retrograde transport pathway in live cells. The Journal of Cell Biology 147 (4), 743–760.
Zhang, B., Enninga, J., Olivo-Marin, J.-C., Zimmer, C., Apr. 2006. Auto-

mated super-resolution detection of fluorescent rods in 2D. In: Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'2006). Washington, DC, pp. 1296–1299.

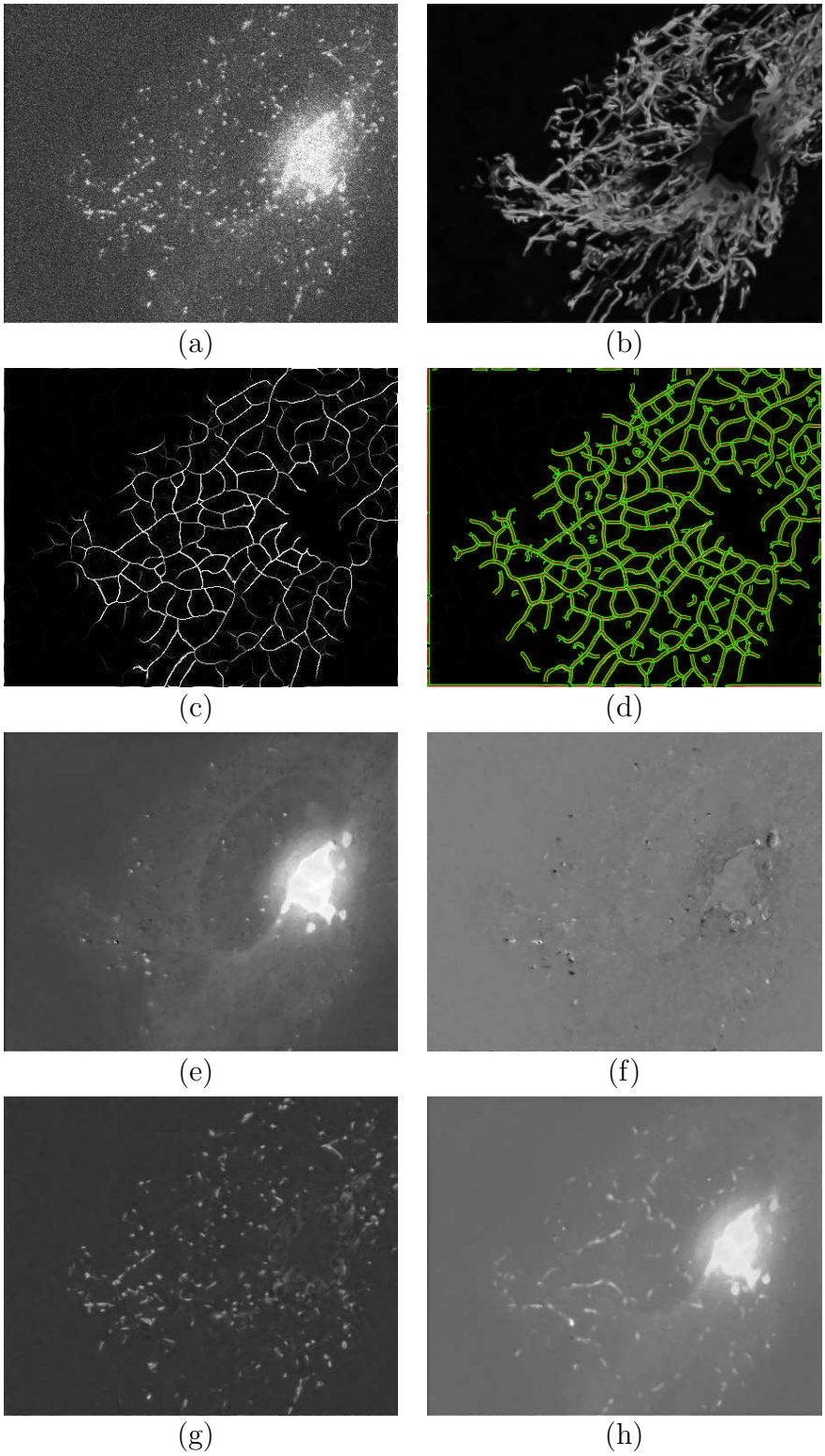

Fig. 9. Simulation of a video-microscopy image sequence exploiting a 3D real images acquired with time-lapse wide-field microscopy. (a) one frame of the sequence projected (maximum) along z axis (depth) computed from the original 3D+time image sequence ; (b) maximum intensity projection 2D map *wrt* time t ; (c) results of steerable filtering ; (d) results of the un-biased line detector ; (e) map **a** ; (f) map **b** ; (g) residual map ; (h) noise-free image reconstruction from estimated parameters.

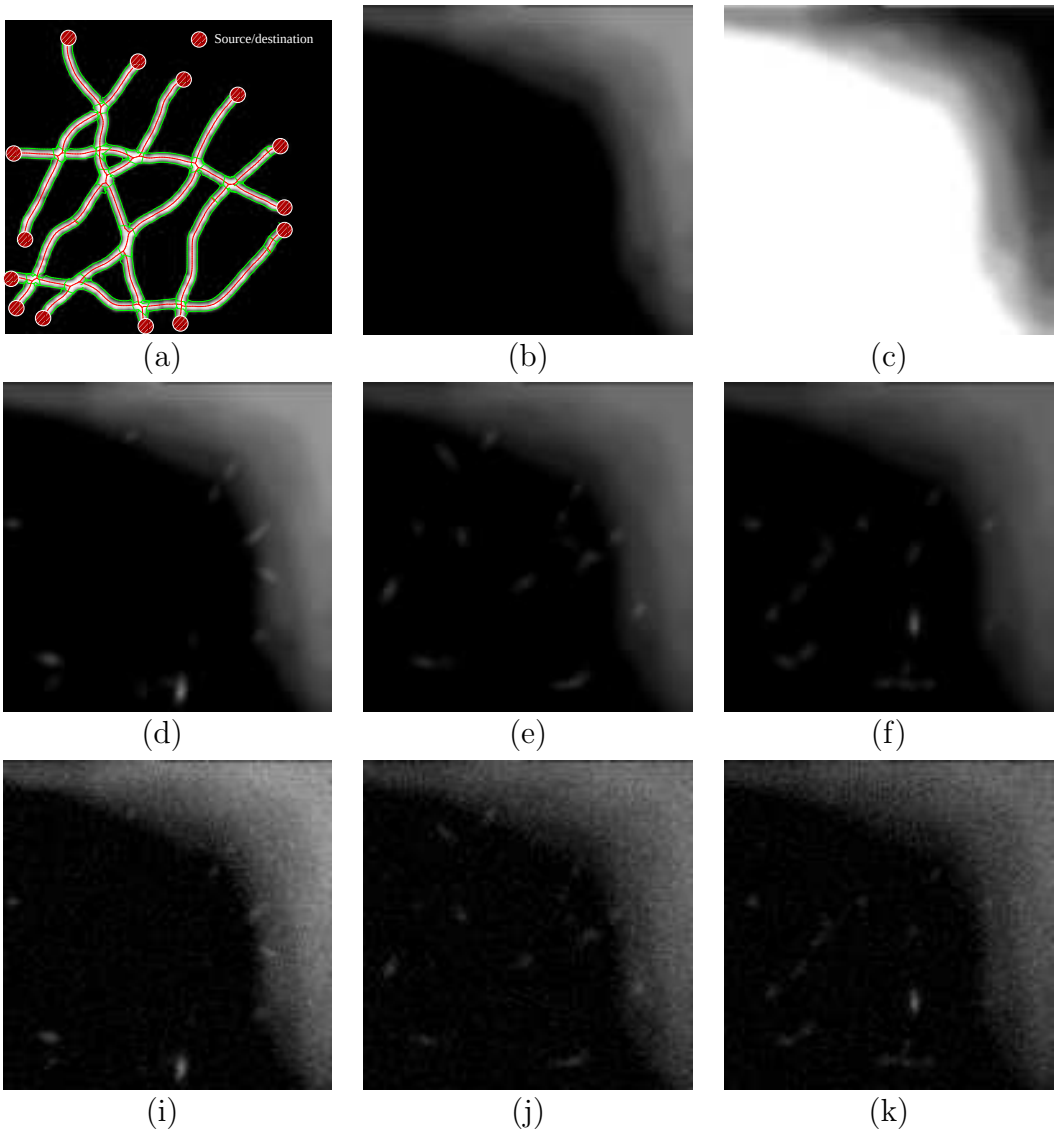

Fig. 10. Simulation of a synthetic video-microscopy image sequence: (a) hand-made network ; (b) true map **a** corresponding to the background model ; (c) true map **b** corresponding to the temporal variation of the background ; (d,e,f) three frames ($t = 0, t = 75, t = 100$) extracted from the noise-free synthetic image sequence ; (i,j,k) three noisy frames ($t = 0, t = 75, t = 100$) corresponding to a signal-to-noise ratio of 13.6dB.

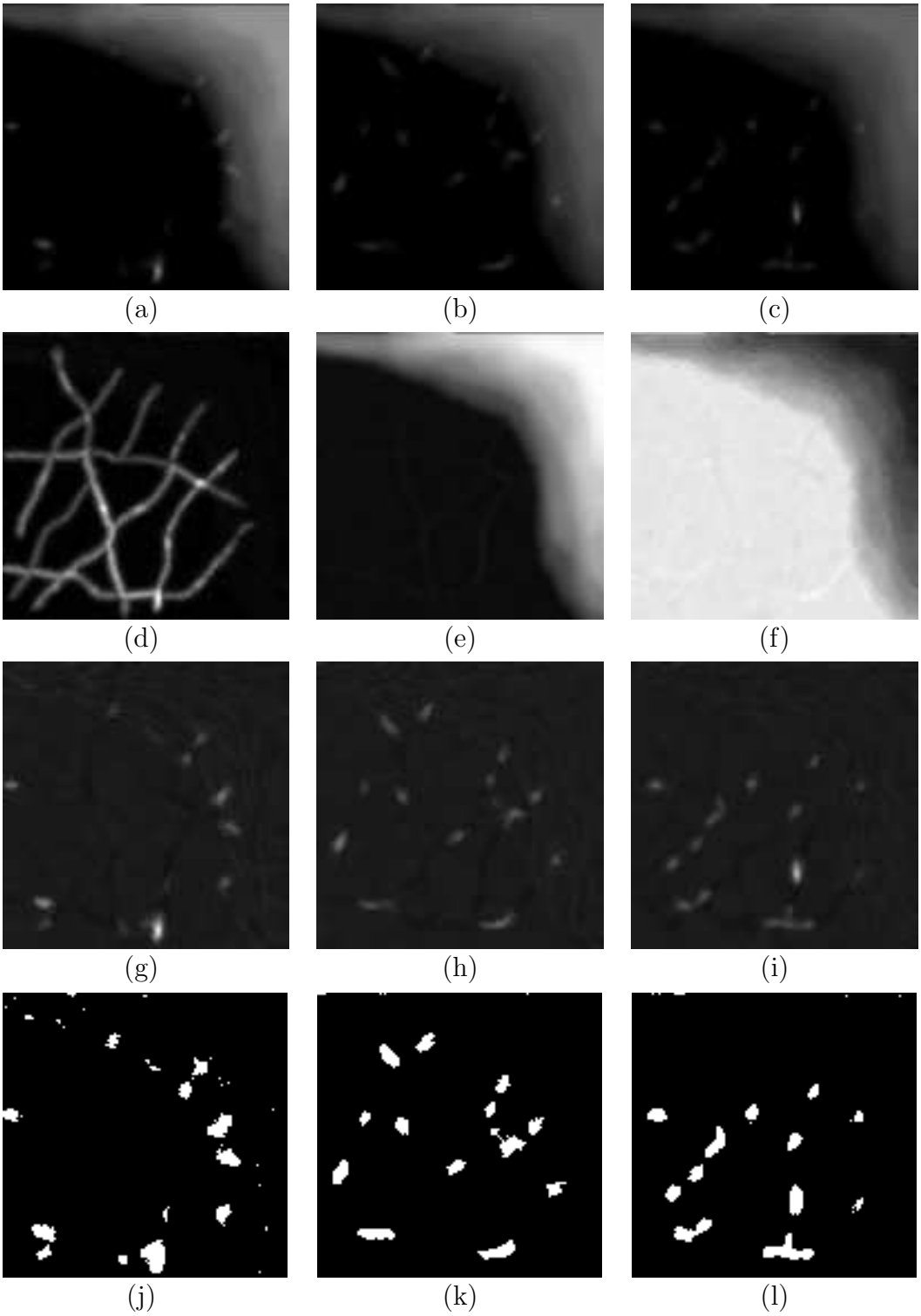

Fig. 11. Computation of the background model parameters and detection of moving noisy vesicles: (a,b,c) three frames ($t = 0, t = 75, t = 100$) of the denoised image sequence ; (d) maximum intensity projection map ; (e) estimated map $\widehat{\mathbf{a}}$; (f) estimated map $\widehat{\mathbf{b}}$; (g,h,i) residual images at time $t = 0, t = 75$ and $t = 100$ (after background subtraction) ; (j,k,l) detected vesicles using the proposed temporal detection method ($t = 0, t = 75, t = 100$).