



HAL
open science

Detection and characterization by local statistical approaches of dynamical events in image sequences : application to membrane fusion in TIRF microscopy

Antoine Basset

► **To cite this version:**

Antoine Basset. Detection and characterization by local statistical approaches of dynamical events in image sequences : application to membrane fusion in TIRF microscopy. Image Processing [eess.IV]. Université Rennes 1, 2015. English. NNT : 2015REN1S096 . tel-01304780

HAL Id: tel-01304780

<https://theses.hal.science/tel-01304780v1>

Submitted on 20 Apr 2016

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.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Traitement du Signal et Télécommunications

Ecole doctorale MATISSE

présentée par

Antoine Basset

Préparée au Centre Inria Rennes – Bretagne Atlantique

**Détection et caractérisation
par des approches
statistiques locales
d'évènements dynamiques
dans des séquences
d'images : Application
à la fusion membranaire
en microscopie TIRF**

**Thèse soutenue à Rennes
le 21 décembre 2015**

devant le jury composé de :

Jean-Marc ODOBEZ
Senior Researcher EPFL / *rapporteur*

Ivo SBALZARINI
Professeur Université de Technologie de Dresde / *rapporteur*

Christine GUILLEMOT
Directrice de Recherche Inria / *examinatrice*

David HOLCMAN
Directeur de Recherche CNRS / *examineur*

Jean SALAMERO
Directeur de Recherche CNRS / *examineur*

Alain TRUBUIL
Ingénieur de Recherche INRA / *examineur*

Patrick BOUTHEMY
Directeur de Recherche Inria / *directeur de thèse*

Charles KERVRANN
Directeur de Recherche Inria / *co-directeur de thèse*

*Essentially, all models are wrong,
but some are useful.*

GEORGES BOX, 1987

Remerciements

MERCI à mes directeurs, Patrick Bouthemy et Charles Kervrann, pour leur présence indéfectible, leur écoute et leurs conseils avisés. À Jean, Jérôme, François, Perrine, pour leurs conseils tout autant avisés, et par dessus tout leur persévérance à expliquer la biologie et l'optique à nos esprits hermétiques d'informaticiens.

À tous les membres de la future équipe commune qui en façonnent déjà l'ambiance chaleureuse. En particulier, merci aux doctorants et stagiaires qui ont accepté de s'impliquer dans mes projets : Juan, Vincent, Anca, Frédéric. Merci à mes collègues étrangers, qui ont travaillé sans relâche à mon « perfectionnement » en mandarin, allemand, roumain, espagnol, arabe et belgicisms...

Merci à mes rapporteurs, Jean-Marc Odobez et Ivo Sbalzarini, de même qu'à François Brémond et Jean-Christophe Olivo-Marin, qui ont accepté et assumé la lourde tâche d'évaluer ma thèse. Aux membres du jury, qui se sont déplacés un 21 décembre, c'est-à-dire probablement au beau milieu de leurs vacances.

Merci à mes proches, qui ont su m'entourer durant ces années, me soutenir et me remotiver quand il le fallait. À ma famille et belle-famille ; en particulier à ma mère, graphiste de son talent, sans qui ce manuscrit ressemblerait à un article IEEE. Pour le soutien moral et tout le reste, merci à Corinne, Jean-Marc, Brigitte, Alex et sa tribu. Évidemment, merci à C.Gaëlle, qui a passé un temps infini à « colorier des petits bons-hommes en bleu » armée d'une patience dépassant l'entendement, ainsi qu'à relire et corriger sans cesse mes articles et ce document.

Merci à tous les courageux qui ont fait le déplacement, parfois long, pour assister à ma soutenance. Un immense merci à Cédric et Angélique, qui nous ont ouvert leurs portes et leurs bouteilles à la veille de cet évènement.

Merci encore et enfin à Gaëlle et à Hélène, largement responsables de mon embauche au CNES, où je m'épanouis désormais au service de la cosmologie.



Évidemment, merci à la belle Région Bretagne et Inria d'avoir financé ma thèse.

Résumé en français

NOTRE TRAVAIL de thèse porte sur la détection et la modélisation de configurations dynamiques dans des séquences d'images en recourant à des approches statistiques locales sans apprentissage supervisé. Deux cas peuvent se présenter :

1. les objets étudiés n'interagissent pas, et les dynamiques *individuelles* peuvent être analysées indépendamment ;
2. les objets étudiés interagissent, et la dynamique à analyser est celle du *groupe* d'objets entier.

En ce qui concerne les dynamiques individuelles, nous nous intéressons à certaines structures de la cellule, les vésicules, qui sont observées à la frontière de la cellule – où elles jouent un rôle essentiel pour la vie et les échanges cellulaires – en microscopie de fluorescence par réflexion totale interne (TIRF).

La dynamique de groupe est notamment rencontrée dans les mouvements de tissus cellulaires, le développement embryonnaire ou les mouvements de foules. Ces derniers sont particulièrement intéressants en termes d'évaluation car de nombreuses méthodes d'analyse de mouvement de foule ont été proposées dans la littérature. Ce domaine d'application représente donc un excellent support expérimental.

Dans les deux cas d'étude, nous nous intéressons aux mouvements de collections d'objets similaires – molécules, cellules ou piétons – évoluant sur un fond immobile, et imagés par un dispositif statique. Du point de vue du traitement d'image, de nombreuses questions sont abordées, notamment la détection (spatio)temporelle d'objets ou d'évènements, la caractérisation de dynamiques ou encore l'estimation de grandeurs physiques. Dans l'un ou l'autre cas d'étude, nous abordons les problématiques selon une démarche commune, essentiellement dirigée par les données et mettant en œuvre des tests statistiques. Par ailleurs, nous avons le souci de proposer des méthodes nécessitant le réglage d'un faible nombre de paramètres, soit peu sensibles, soit calibrés avec des règles statistiques.

Nous nous intéressons en premier lieu à l'analyse de dynamiques membranaires de la cellule en microscopie optique de fluorescence. Pour cause, la vie de la cellule est sujette à un équilibre très précis des différents éléments chimiques qui la peuplent. La régulation de cet équilibre passe nécessairement par des échanges avec le milieu extérieur, donc par le franchissement de la membrane plasmique qui encoint la cel-

lule. Notre objectif est de caractériser les dynamiques de vésicules évoluant dans la cellule. Les vésicules sont des intermédiaires de transport véhiculant divers composants dans travers la cellule. Dans le cas de l'exocytose, les vésicules déplacent ces composants à l'extérieur de la cellule. Pour ce faire, les composants sont transportés jusqu'à la membrane plasmique de la cellule, puis la vésicule fusionne avec celle-ci pour libérer son contenu.

Pour analyser ces dynamiques membranaires, un premier travail consiste à détecter les vésicules d'intérêt. Pour cela, nous proposons une méthode de détection de spots performante et simple à paramétrer. Une analyse, requérant des informations plus biologiques et biophysiques, permet ensuite de détecter et reconnaître certains événements dynamiques dans la cellule, en particulier la fusion de la vésicule avec la membrane plasmique. Les paramètres biophysiques associés sont ensuite estimés. Dans ce cadre, nous proposons des modèles de fusion, dont nous estimons les paramètres pour plusieurs conditions expérimentales. En particulier, nous mettons en évidence les comportements différents de deux protéines d'intérêt : le récepteur à la transferrine (TfR) et la Langérine.

Dans la troisième partie de la thèse, nous étudions la classification de mouvements de groupes, la détection des chemins les plus empruntés dans la vidéo et la détection de comportements anormaux rares, parfois suspects. Comme indiqué plus haut, nous avons principalement évalué nos méthodes d'analyse de mouvements de groupe sur des vidéos de foules. Aucune comparaison de méthode n'est actuellement possible sur des images de biologie. En revanche, nos approches trouvent un intérêt certain pour analyser des comportements collectifs souvent observés en imagerie biologique.

La thèse comprend trois parties : la détection de spots, l'analyse de dynamiques individuelles et l'analyse de dynamiques de groupes.

Partie I. Détection de spots

Dans le cadre de l'analyse d'images de microscopie, la détection automatisée des éléments à étudier constitue la plupart du temps une étape préliminaire essentielle qui conditionne l'ensemble des analyses ultérieures, que ce soit le suivi de vésicules [Meijering et al., 2012], la classification de mouvements [Sage et al., 2005] ou la caractérisation de dynamiques membranaires [Mele et al., 2009]. Dans ce dernier cas, nous cherchons à segmenter les vésicules imagées en microscopie de fluorescence à réflexion totale interne (TIRF). Ces vésicules sont généralement de tailles semblables, et il convient donc d'estimer l'échelle caractéristique de ces objets.

Des études comparatives [Smal et al., 2010, Rezatofghi et al., 2012] ont permis d'évaluer les performances de nombreuses méthodes de détection non supervi-

sées : produit multi-échelle d'ondelettes (WMP [Olivo-Marin, 2002]), détection multi-échelle après stabilisation de variance (MS-VST [Zhang et al., 2007]), détection de « dômes » (HD [Smal et al., 2008], MPHD [Rezatofghi et al., 2012]), filtre de rehaussement de contraste de spots (SEF [Sage et al., 2005], TH [Bright and Steel, 1987], MTH [Soille, 2003]), extraction de points caractéristiques (IFD [ter Haar Romeny, 2003])...

La plupart de ces méthodes nécessitent un ajustement fin de plusieurs paramètres pour obtenir de bons résultats [Rezatofghi et al., 2012, Smal et al., 2010]. Plus précisément, ces détecteurs ont en commun deux types de paramètres : l'*échelle* des objets à détecter et le *seuil* de détection. Afin de réduire le nombre et la sensibilité des paramètres, nous développons une méthode de segmentation de vésicules avec sélection d'échelle automatique et seuillage adaptatif, dénommée ATLAS¹. Elle s'appuie sur un filtre laplacien de gaussienne (LoG), dont la variance est automatiquement sélectionnée dans un ensemble fini de valeurs prédéfinies. Un seuillage adaptatif local permet alors de décider si un pixel appartient ou non à une vésicule, même lorsque l'arrière-plan est très complexe. Sous l'hypothèse d'une distribution localement gaussienne des intensités de l'image, le seuil est inféré, en chaque pixel, d'une probabilité de fausse alarme choisie par l'utilisateur pour l'ensemble des images traitées.

Sélection automatique de l'échelle

La sélection d'échelle s'appuie sur une représentation multi-échelle de l'image I à segmenter. La représentation multi-échelle est une famille d'images $\{L_s\}_{s \in \mathcal{S}}$ obtenues par convolution avec des noyaux G_s d'échelles s croissantes prises dans un ensemble de valeurs prédéfinies $\mathcal{S} \subset \mathbb{R}_+^*$:

$$\forall s \in \mathcal{S}, L_s = G_s * I,$$

où $*$ désigne l'opération de convolution. Pour construire la représentation multi-échelle d'un signal continu, on peut utiliser des noyaux gaussiens de variance s [Lindeberg, 1998]. Cependant, pour des signaux discrets comme les images, il est recommandé d'utiliser des noyaux également discrets analogues au noyau gaussien, appelés plus simplement *noyaux gaussiens discrets*. On les obtient à l'aide des fonctions de Bessel B_n [Lindeberg, 1990] :

$$\forall s \in \mathcal{S}, \forall n \in \mathbb{Z}, G_s(n) = e^{-s} B_n(s).$$

On applique ensuite un opérateur laplacien normalisé noté $\alpha(s)\Delta$ qui permet d'augmenter le contraste des vésicules par rapport à l'arrière-plan [Sage et al., 2005]. Le scalaire $\alpha(s)$ désigne un facteur de normalisation de G_s . Par associativité et commutativité de l'opération de convolution, le laplacien peut être appliqué une seule fois

¹Adaptive threshold of LoG with auto-selected scale

directement sur I , et non pas à chaque échelle. On obtient ainsi le *LoG multiéchelle* H :

$$\forall s \in \mathcal{S}, H_s = \alpha(s) G_s * \Delta I.$$

Nous cherchons ensuite, parmi les échelles de \mathcal{S} , celle qui permet d'augmenter au mieux le contraste des vésicules. Pour la déterminer, on s'intéresse aux minima locaux de H dans l'espace $\mathcal{S} \times \Omega_I$, appelés *blobs*. Ces blobs sont principalement localisés en deux lieux précis :

- au centre des spots gaussiens correspondant aux vésicules ;
- au niveau des pixels brillants induis par le bruit.

Nous cherchons l'échelle pour laquelle la plus grande proportion de blobs est issue des vésicules, et non du bruit. La distribution du nombre de blobs par unité de surface en fonction de l'échelle nous informe sur l'échelle optimale. En effet, on peut comparer cette distribution calculée sur l'image I à la distribution obtenue sur une image g ne présentant que du bruit. Les courbes obtenues pour I et g se distinguent autour d'une certaine échelle s^* , caractéristique de l'image I , qui offrira les meilleurs résultats de détection. Pour la détecter, nous proposons de maximiser le rapport entre les deux distributions. Ceci revient à faire un compromis entre taux de bonnes détections (spots) et taux de fausses détections (bruit).

Segmentation adaptative des spots

Une fois l'échelle optimale s^* sélectionnée, nous nous concentrons sur le LoG correspondant, H_{s^*} , que nous seuillons pour ne retenir que les valeurs les plus faibles, correspondant principalement aux vésicules.

Lorsque l'arrière-plan est complexe et le contraste local variable, un seuil global ne permet pas de segmenter correctement les vésicules. Il faut alors adapter la valeur du seuil au contenu local de l'image. Pour cela, nous proposons de déterminer ce seuil, en chaque point, en considérant l'histogramme local du LoG. Nous calculons d'abord, en tout point $p \in \Omega_I$, la moyenne locale $\mu(p)$ et la variance locale $\sigma^2(p)$ de H_{s^*} sur une fenêtre W :

$$\begin{aligned} \mu(p) &= (W * H_{s^*})(p), \\ \sigma^2(p) &= (W * H_{s^*}^2)(p) - \mu^2(p). \end{aligned}$$

En supposant que H_{s^*} a une distribution locale gaussienne, on peut alors déduire le seuil local $\tau(p)$:

$$\tau(p) = \Phi^{-1}(P_{\text{FA}}) \sigma(p) + \mu(p),$$

où Φ est la fonction de répartition de la loi normale centrée réduite et P_{FA} est une probabilité de fausse alarme choisie par l'utilisateur.

Résultats expérimentaux

Nous utilisons différents bancs de test introduits dans [Smal et al., 2010, Pécot et al., 2015, Ruusu vuori et al., 2012], et nous en proposons quatrième², pour comparer notre méthode aux détecteurs de l'état de l'art : WMP [Olivo-Marin, 2002], MS-VST [Zhang et al., 2007], TH [Bright and Steel, 1987], MTH [Soille, 2003], HD [Smal et al., 2008], MPHD [Rezatofghi et al., 2012], SEF [Sage et al., 2005], IDF [ter Haar Romeny, 2003], C-CRAFT [Pécot et al., 2015] et LR-MRF [Ruusu vuori et al., 2012]. Dans la quasi-totalité des cas, nous obtenons un meilleur score que la méthode la plus performante.

Qualitativement, nous mettons en évidence que la précision de la segmentation des spots sur images réelles, en particulier en comparant nos cartes de détection à celles de la méthode obtenant les meilleurs résultats sur simulations après ATLAS, à savoir MS-VST.

Le logiciel ATLAS peut être testé en ligne sur le portail web de l'équipe-projet Serpico³.

Partie II. Analyse des dynamiques individuelles

Cette partie de la thèse est menée en collaboration avec l'équipe de Jean Salamero à l'Institut Curie, UMR 144 CNRS.

Nous étudions des complexes moléculaires impliqués dans les dernières étapes de l'exocytose, à savoir la fin de la phase de transport par les vésicules (le long des microtubules, puis dans le réseau de filaments d'actine), l'arrimage à la membrane et la fusion membranaire. Nous nous intéressons à trois protéines associées aux vésicules : Rab11, associée aux vésicules durant la phase de transport, et Langérine et TfR, deux protéines transmembranaires transportées par des vésicules.

Pour imager la cellule vivante, on utilise généralement en microscopie optique des techniques de microscopie de fluorescence. Un faisceau laser est émis en direction de la cellule pour exciter des marqueurs fluorescents qui ont été liés aux protéines d'intérêt (pHluorine pour Langérine et TfR, mCherry pour Rab11). Dans notre étude, nous utilisons la microscopie TIRE, qui permet de n'exciter qu'une épaisseur très faible de la cellule au niveau de la membrane, et ainsi d'isoler la région où ont lieu les dernières étapes de l'exocytose [Axelrod, 2008]. En pratique, le faisceau laser est émis avec un angle d'incidence supérieur à l'angle limite de réfraction, de sorte qu'il est totalement réfléchi par l'interface entre la lamelle de verre et la cellule. Cependant, une faible quantité d'énergie (l'onde évanescente) pénètre dans la cellule, et son amplitude décroît exponentiellement avec la profondeur. Ainsi, seules les protéines marquées proches de l'interface sont principalement excitées.

²serpico.rennes.inria.fr

³mobyli-serpico.rennes.inria.fr

Classification des dynamiques membranaires de Rab11

Notre objectif est de caractériser le comportement dynamique des vésicules et des molécules cargo, c'est-à-dire transportées par la vésicule, au moment de l'exocytose. Dans un premier temps, il faut détecter le lieu et le moment où se produit l'exocytose. Nous proposons une méthode de classification instantanée simple s'appuyant sur un test statistique d'hypothèses. Ces hypothèses sont élaborées à partir d'une représentation mathématique du mécanisme d'exocytose.

Avant d'atteindre la membrane puis de libérer leur contenu, les vésicules transistent sur le cytosquelette. Certaines vésicules s'amarrent alors à la membrane et peuvent ensuite fusionner avec elle. Notons I la séquence d'images à analyser et $I(p, t)$ la valeur au point p dans la t -ième image. Nous proposons deux modèles dynamiques pour décrire l'évolution spatio-temporelle des vésicules :

- *Translation* pour les vésicules en phase de transport ou d'amarrage, avec conservation de l'intensité :

$$\frac{\partial I}{\partial t}(p, t) = -\nabla I(p, t) \cdot \mathbf{w},$$

où $\mathbf{w} = (u, v)^T$ est la translation de la vésicule segmentée, valable pour tous les points de la vésicule, et ∇ désigne l'opérateur gradient ;

- *Diffusion* pour les vésicules fusionnant avec la membrane :

$$\frac{\partial I}{\partial t}(p, t) = D \Delta I(p, t),$$

où D est le coefficient de diffusion pour cette vésicule et Δ désigne l'opérateur laplacien 2D.

Ces modèles dynamiques ne sont valides que pour des vésicules présentant un changement de fluorescence dans le temps. Une étape préalable à cette classification consiste à détecter, parmi les vésicules segmentées par ATLAS, celles qui se déplacent ou changent d'apparence, vésicules que nous qualifierons de *dynamiques*. Une vésicule segmentée à l'instant t est considérée comme dynamique si au moins un de ses pixels subit un changement en t [Boulanger et al., 2010a].

Une fois les vésicules dynamiques détectées, nous estimons les paramètres des modèles (respectivement, les deux composantes de translation u, v et le coefficient de diffusion D) par une technique de moindres carrés. Pour distinguer l'état dynamique (translation ou diffusion) de chaque vésicule, nous procédons à un test de vraisemblance généralisé pour comparer les deux modèles ainsi estimés et retenons le modèle qui a la meilleure vraisemblance.

Modèle de fusion pour TfR et Langérine

Pour comparer les comportements de TfR et Langérine, deux protéines transmembranaires, nous proposons un modèle de diffusion augmenté d'un paramètre, dit *temps*

de résidence, τ , qui coïncide avec la durée nécessaire à une protéine pour quitter la vésicule :

$$I(p, t) = \frac{A_0/\tau}{\sigma_{\text{PSF}}^2} \exp\left(-\frac{t}{\tau} - \frac{\|p\|_2^2}{2\sigma_{\text{PSF}}^2}\right) + \int_0^t \frac{A_0/\tau}{2D(t-u) + \sigma_{\text{PSF}}^2} \exp\left(-\frac{u-t_0}{\tau} - \frac{\|p\|_2^2}{4D(t-u) + 2\sigma_{\text{PSF}}^2}\right) du,$$

où A_0 est l'amplitude initiale du spot et σ_{PSF}^2 sa variance initiale.

Les évènements de fusion membranaire sont plus aisément détectés pour TfR et Langérine que pour Rab11. Ces protéines transmembranaires sont en effet associées à un marqueur pH-sensible. Comme le pH augmente subitement au moment de la fusion, nous détectons fiablement les spots apparaissant au fil de la séquence, toujours à l'aide d'ATLAS. Parmi les évènements de fusion, on observe que certains ne donnent pas lieu à une diffusion. Pour éliminer ceux-ci, nous proposons un test statistique d'adéquation au modèle de fusion.

Une fois les évènements diffusifs détectés, les paramètres du modèle de diffusion sont également estimés par une technique de moindre carrés, mais certaines améliorations sont proposées pour obtenir des résultats performants. En particulier, nous mettons en concurrence plusieurs estimations obtenues avec plusieurs initialisations, et nous retenons celle qui minimise la somme des carrés des résidus.

Des expériences quantitatives sur simulation mettent en évidence les performances de notre méthode d'estimation. Nous traitons par la suite un grand jeu de séquences réelles, pour lesquelles est marquée la TfR ou la Langérine. Ceci nous permet finalement d'exhiber les différences de comportement de ces deux protéines au moment de la fusion, en comparant les distributions du temps de résidence τ dans ces deux cas.

Partie III. Analyse des dynamiques de groupes

Dans cette dernière partie, nous nous intéressons au mouvement collectif d'entités mobiles similaires. Comme nous l'avons expliqué en introduction, même si les méthodes sont applicables à des images biologiques, nous utilisons des vidéos de foule comme support de validation expérimentale.

La vidéosurveillance génère d'énormes quantités de données, qui ne peuvent pas être intégralement visionnées, et nécessitent donc des outils d'analyse automatisés [Zhan et al., 2008]. En particulier, l'analyse des mouvements de foule dans les vidéos est un domaine extrêmement vaste et en pleine expansion, qui englobe notamment la sécurisation des évènements publics, la surveillance des lieux des transports en commun et l'analyse des comportements individuels au sein de groupes.

Dans ce contexte, nous abordons trois aspects importants :

- la classification de mouvements de groupes [Garate et al., 2014, Solmaz et al., 2012, Wang et al., 2011, Wu et al., 2014, Zhou et al., 2013],
- la recherche des chemins les plus empruntés [Wang et al., 2011, Zhou et al., 2011],
- la détection de comportements anormaux [Benezeth et al., 2011, Kim and Grauman, 2009, Kratz and Nishino, 2009, Mehran et al., 2009].

Pour répondre à ces problématiques, nous proposons une nouvelle approche reposant sur l'analyse d'une paire ou un triplet d'images seulement, alors que les méthodes usuelles d'analyse de foules requièrent des intervalles de temps beaucoup plus longs (au moins une dizaine d'images, et parfois même des minutes de vidéo). Ces approches exploitent en général des cuboïdes spatiotemporels [Feng et al., 2010, Kratz and Nishino, 2009, Rodriguez et al., 2011], des « tracklets » [Garate et al., 2014, Zhou et al., 2011] ou plus généralement des trajectoires [Cheriyadat and Radke, 2008, Rodriguez et al., 2009, Solmaz et al., 2012, Wang et al., 2011, Zhou et al., 2012]. Contrairement à ces méthodes, les nôtres reposent sur la mesure de champs affines de mouvement 2D entre images successives, calculées sur une collection de fenêtres : nous parlerons d'approche *instantanée*. Nous évitons ainsi d'une étape coûteuse et souvent difficile de suivi individuel des piétons [Baumgartner et al., 2013, Cupillard et al., 2002, Hu et al., 2008, Idrees et al., 2014, Kratz and Nishino, 2012, Rodriguez et al., 2009] ou d'advection de particules à partir d'un flot optique [Solmaz et al., 2012, Wu et al., 2010]. Dans nos méthodes, nous n'utilisons ni intégration temporelle, ni calcul de trajectoire, ni phase d'apprentissage.

Sélection ponctuelle de types de mouvement 2D

Nous abordons les trois problématiques sus-citées avec un descripteur commun de mouvement, basé sur une sélection ponctuelle de types de mouvements affines 2D. Notre approche n'exploite pas de modèle de comportement humain, et peut s'étendre à l'analyse de groupes d'entités en mouvement : véhicules, animaux, entités biologiques...

Soit $\mathcal{W} = \{W_i\}_{i \in \mathbb{N}}$ une collection de fenêtres de tailles variées et se chevauchant. Nous considérons trois types de mouvements affines 2D, respectivement *translation* (T), *divergence* (D) et *rotation* (R). Nous notons $\theta_{k,i}$ les paramètres du modèle $k \in \{T, D, R\}$ pour la fenêtre W_i . En chaque point $p = (x, y)$ de W_i et pour chaque modèle $k \in \{T, D, R\}$, la vitesse $w_{k,i}$ est donnée par les formules suivantes :

- Pour la translation :

$$w_{T,i} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \text{avec} \quad \theta_{T,i} = (b_1, b_2)^T;$$

- Pour la divergence :

$$\mathbf{w}_{D,i} = \begin{pmatrix} a_1 & 0 \\ 0 & a_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \text{avec} \quad \boldsymbol{\theta}_{D,i} = (a_1, b_1, b_2)^T;$$

- Pour la rotation :

$$\mathbf{w}_{R,i} = \begin{pmatrix} 0 & -a_2 \\ a_2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \text{avec} \quad \boldsymbol{\theta}_{R,i} = (a_2, b_1, b_2)^T.$$

Cette façon de mesurer le mouvement permet d'éviter de calculer le flot optique, et facilitera la classification ultérieure du mouvement.

Les paramètres des modèles sont estimés par une méthode robuste dans un schéma d'estimation multirésolution [Odobez and Bouthemy, 1995] avec le logiciel Motion2D⁴. Nous obtenons ainsi de l'ordre d'une centaine de vecteurs de vitesse candidats par point (trois par fenêtre dans une trentaine de fenêtres). Nous proposons de sélectionner le meilleur candidat selon un critère de maximum de vraisemblance pénalisé basé sur l'hypothèse de conservation d'intensité. La conformité d'un point p au modèle de vitesse de paramètre $\boldsymbol{\theta}_{k,i}$ est représentée par la variable aléatoire $\varepsilon_{k,i}$ définie par :

$$\varepsilon_{k,i}(p) = I(p + \mathbf{w}_{k,i}, t + 1) - I(p, t).$$

En supposant que chaque variable suit une loi gaussienne centrée, et qu'elles sont indépendantes, on peut calculer la vraisemblance jointe dans un patch centré en p . Le meilleur candidat en p est ensuite sélectionné selon le critère d'information d'Akaike corrigé pour tenir compte du petit nombre d'observations fournies par le voisinage local du point p [Burnham and Anderson, 2002].

Classification de mouvements d'ensemble

Des trois types de mouvement affine considérés, nous déduisons huit classes de mouvement de groupe en fonction des valeurs des paramètres estimés. Les mouvements divergents correspondent au rapprochement du groupe (*divergence positive*) ou à son éloignement (*divergence négative*). Nous distinguons par ailleurs les mouvements de rotation *directe* et *indirecte*, et quatre directions principales de translation pertinentes dans le référentiel de l'image : *nord*, *ouest*, *sud* et *est*. Nous obtenons ainsi une carte de classification préliminaire \hat{c} , forcément bruitée.

Pour obtenir la classification finale C des mouvements de foule, nous régularisons \hat{c} par une méthode de votes majoritaires. De plus, on voit souvent apparaître côte-à-côte dans \hat{c} des mouvements de divergences ou de rotations opposés. Pour contourner cette difficulté, nous introduisons un processus de vote en deux tours. Au premier

⁴www.irisa.fr/vista/Motion2D

tour, nous introduisons des interactions entre classes, appelées *inhibition* et *renforcement* ; il permet de sélectionner le type de mouvement : translation, divergence ou rotation. Le second tour permet d'affiner la classification pour obtenir la carte finale régularisée avec huit classes de mouvement de foule.

Détection des chemins principaux

Connaissant le type de mouvement affine sélectionné en chaque point, nous déduisons le vecteur de vitesse en chaque point. Nous obtenons ainsi une approximation pertinente du flot optique. Nous estimons alors la vitesse moyenne dans chaque cellule d'une grille régulière et sur un court intervalle de temps. Cette mesure suppose un mouvement stationnaire dans la vidéo traitée, hypothèse cohérente avec la recherche des chemins les plus empruntés dans la scène. Pour retrouver le chemin le plus emprunté, nous discrétisons les directions de ces vitesses moyennes (en huit secteurs) de façon à pointer vers une cellule voisine.

Nous obtenons ainsi un graphe orienté où les nœuds correspondent aux cellules de la grille et les arêtes sont définies par les directions locales du flot optique moyen. Ce graphe est en fait réduit à un arbre, car de chaque nœud ne part qu'une arête au plus. La recherche du plus long chemin dans ce graphe est donc très simple (il suffit de vérifier qu'il n'existe pas de cycle). Sous l'hypothèse d'un mouvement stationnaire, nous détectons ainsi les chemins les plus suivis dans la scène.

Détection de comportements anormaux

A partir de la carte de classification préliminaire \hat{c} , nous pouvons également détecter et localiser les comportements localement anormaux. Un mouvement est considéré comme anormal s'il diffère significativement du mouvement environnant.

Nous introduisons une comparaison d'histogrammes des classes locales de mouvement de foule. En chaque point p , on mesure la distribution des huit classes de mouvement dans un voisinage $\eta(p)$ de ce point. Nous comparons cette distribution empirique à celles de huit patchs voisins $\eta_i(p)$ répartis autour de p . Une mesure de distance entre histogrammes est calculée, prenant en compte la spécificité des classes. Les translations ont un caractère circulaire, tandis que les divergences, comme les rotations, s'opposent entre elles.

Nous considérons qu'une anomalie, associée à un mouvement local, peut être caractérisée en chaque point p à travers la distance minimale entre l'histogramme des classes dans le patch central $\eta(p)$ et ceux dans les patchs voisins $\eta_i(p)$. Autrement dit, un mouvement local est anormal s'il ne ressemble à aucun mouvement situé à proximité. Le minimum de ces distances est supposé suivre une distribution des valeurs extrêmes généralisées [Embrechts et al., 1997], dont nous estimons les paramètres

avec la méthode décrite dans [Boulanger et al., 2010a]. Nous déduisons ainsi un seuil sur les distances à partir d'une probabilité de fausse alarme choisie par l'utilisateur, qui contrôle précisément le taux de fausses détections.

Les expériences sur des séquences simulées et réelles mettent en évidence quantitativement l'efficacité de la méthode, tant en termes de détection que de localisation des anomalies.

Contents

Résumé en français	vii
Partie I. Détection de spots	viii
Partie II. Analyse des dynamiques individuelles	xi
Partie III. Analyse des dynamiques de groupes	xiii
1 General introduction	3
1.1 Context and motivation	4
1.2 Objectives and contributions	7
1.3 Organization of the thesis	9
2 On the cell and its observation	13
2.1 A bit of history	14
2.2 Organization and processes of the cell	16
2.3 TIRFM for imaging membrane dynamics	19
I Spot detection	27
3 The landscape of spot detection	29
Related work	31
4 Automatic selection of the detection scale	37
4.1 Scale-space representations	37
4.2 Blob distribution and scale selection	40
5 Spot segmentation	47
5.1 Adaptive thresholding	47
5.2 Estimation window	48
5.3 Experimental results	50
5.4 Discussion	64

II Individual dynamics	69
6 Modeling of the vesicle fusion dynamics	71
6.1 Exocytosis process	72
6.2 Existing diffusion models	76
6.3 Proposed fusion models for Rab11	84
6.4 Proposed SSED model for TfR and Langerin	88
6.5 Discussion	92
7 Detection of diffusion dynamics	95
7.1 Detection based on Rab11 dynamics classification	96
7.2 Event detection using fluorophore pH-sensitivity	102
7.3 Discussion	107
8 Estimation of the biological parameters	109
8.1 Existing diffusion estimation methods	110
8.2 Proposed estimation method for the translation	114
8.3 Proposed estimation method for the point source model	114
8.4 Proposed estimation method for the SSED model	118
8.5 Discussion and perspectives	121
III Group dynamics	125
9 Labeled affine flow for group motion characterization	127
9.1 Related work	129
9.2 Labeled affine flow	131
9.3 Motion patterns recognition	135
9.4 Recovery of principal paths	139
9.5 Discussion and perspectives	144
10 Anomaly detection and localization	147
10.1 Related work	149
10.2 Histogram distance	151
10.3 Pixel-level anomaly detection and localization	154
10.4 Experimental results	154
10.5 Discussion	160
11 General conclusion	163
11.1 Main contributions and results	164
11.2 Ongoing and future work	165

Bibliography	169
List of publications	189
Acronyms	191

1

General introduction

IN THIS THESIS, we investigate statistical methods to detect, estimate and characterize dynamical events in image sequences. There are two cases:

- Studied objects do not interact, and *individual* dynamics can be independently analyzed;
- Studied objects interact, and *group* dynamics must be analyzed as a whole.

In the case of individual dynamics, our primary focus is on biological image sequences showing proteins evolving in a cell, and more precisely at the cell frontier named plasma membrane. The images are acquired by the means of total internal reflection fluorescence microscopy (TIRFM), an observation technique dedicated to plasma membrane dynamics analysis. We model the protein dynamics and estimate the biophysical parameters in TIRFM image sequences for further biological analysis. We first propose a new spot detection method with automatic scale selection, aimed at localizing dynamical events of interest.

Eventually, we focus on image sequences at lower magnifications, that is, depicting groups of cells, for example cell tissues, instead of an isolated cell. Since they are free of appearance model, the developed methods are quite general and extended to other applications including crowd motion analysis in videos.

Whether it is for spot detection, protein dynamics estimation or group motion analysis, a common philosophy is ubiquitous, however. First, *statistical arguments* are used to automatically infer the method parameters. Therefore, the method does not need to be specified for every single image sequence, and the method can be used easily by neophytes in image processing. Secondly, we rely on *local approaches* which have the advantage of being computationally efficient. Their complexity being typically proportional to the image sequence data volume, they are particularly recommendable in a context of increasingly big data, for which global approaches cannot provide reasonable computation time. Local modeling handles spatially varying im-

age statistics much more easily and more accurately than global modeling, for which the complexity rapidly explodes if one wants to finely model the entire image. Indeed inevitable stationarity assumptions of global approaches are much reduced in local approaches. Moreover, the local characteristics of the approach is also true in the temporal dimension when we process image sequences. Such *instantaneous methods*, using only a couple of consecutive frames to estimate the dynamics, offer a lot of advantages, including the possibility of ensuring real-time computation if needed. Local approaches also allow neglecting low frequency variations such as spatially varying background contrast or, in fluorescence microscopy, temporal fading known as photobleaching.

1.1 Context and motivation

Human body is composed of 10^{13} to 10^{14} cells, millions of which are renewed every second. Correct cell functioning is an essential prerequisite to life, and a segment of biology in which much remains to be discovered. A key actor of life understanding, cell biology lies at the intersection of the very applied and fundamental life sciences, for example as a cornerstone of evolution theory. As such, it also interacts with many research fields, ranging from therapy-related fields (e.g., medicine, drug development, epidemiology) to technology-related ones (e.g., optics, electronics, physics), and relies more and more on computer sciences and applied mathematics, specifically image processing.

In order to detect variations between normal and pathological situations, intra- and inter-cellular dynamics must be understood. This is done through the observation and the characterization of molecular processes, among others, which can be tackled in various ways, from biochemistry to microscopy imaging. We are interested in the latter, which encompasses microscope technology, genetics engineering and processing tools. While optical and biological sciences are out of our reach, we are developing image processing methods, which can profit the biology research.

1.1.1 A need for automatic methods

Joint research efforts in optics, electronics, chemistry and biology, among others, came to tremendous advances in observation technologies, cornerstone of modern biology research. Progress in microscopy imaging make possible the study of cell mechanisms *in vivo* and at sub-micron dimensions. Complex observation techniques can even go further in terms of resolution (e.g., super-resolution methods based on multiple exposures or electron microscopy), but, for now, they do not provide a fine temporal resolution, or do not even allow for acquiring temporal sequences or observing living cells. In contrast, conventional fluorescence microscopy features a very

fast acquisition rate, thanks to a continuous improvement of the sensitivity of digital camera sensors. This enables the use of fluorescent tags which themselves emit the light acquired by the microscope when excited. Researchers in genetics engineering developed protocols to tag the structures of interest, in order to locate and follow the latter in *living* cells, a breakthrough in cell biology. Modern cameras are so sensitive that sub-second exposure times are sufficient to detect and track fluorescent tags, and enable the fine recording of (intra)cellular dynamics. As an order of magnitude, in this thesis, we typically consider 100-nanometer structures traveling 10 microns per second, which represents a few pixels per frame at most.

However, advances in microscopy bring new challenges, both in terms of biological queries and massive data handling. With the improvement of image resolution and with the advances in cell biology knowledge come more specific and complex questions. Expert visual analysis is no more sufficient to reliably answer new queries in biology, while human annotation is not fine enough to deal with the required accuracy. As an example, a classical way of characterizing the actors involved in a biological process is to inhibit the said actors one after the other and see if the process has changed. Changes can be either qualitative – something new or different has happened, or something does not occur anymore – or quantitative – some biological parameter value has changed. While visual inspection is sometimes sufficient to exhibit qualitative changes, parameter variations are generally impossible to assess quantitatively without the help of image processing and statistical tools. Moreover, dramatic improvement of spatial and temporal resolutions yields a so-fast increase of the amount of data to process, and computing methods must now cope with emerging *big big* data at the petabyte scale. The challenge of confronting biological models to such rich observations has yet to be faced. Hence, the need of automatic image processing tools in support of cell biology.

For the three years of the thesis, we have collaborated with Jean Salamero's team at UMR 144 (CNRS, Institut Curie), a team interested in deciphering the dynamics of intracellular transports by the means of spatiotemporal imaging. In response to the need of digital tools, the purpose of the collaboration is to develop methods for the fully automatic analysis of the dynamics of biological structures observed by the means of optical microscopy imaging. To this end, dynamical models representing biological processes are proposed and evaluated on numerous microscope image sequences.

1.1.2 Key role of proteins in the recycling process

Vital processes of the cell, from feeding to motility and reproduction, are subject to the precise equilibrium between the different chemical compounds present in the cell. The equilibrium state is named homeostasis. While from Latin *homeo*, equilibrium,

and *stasis*, static, the term refers to a very dynamical process, or more precisely a set of interacting processes. Absolute and relative quantities of proteins, sugars, ions, among others, are therefore continuously regulated by the cell. The control of the total amount of a given compound is carried out by generating incoming and outgoing flows. Hence, the processes occurring at the cell frontier, the plasma membrane, are of utmost importance to the cell life.

Among various and complex dynamics taking place to ensure homeostasis, exocytosis is the process by which compounds are expelled out of the cell, by crossing the plasma membrane. While smallest structures like ions can autonomously diffuse through the membrane, larger ones rely on transport and exocytosis intermediates called secretory vesicles, or recycling vesicles in our specific case study. To fulfill their function, the vesicles are involved in a variety of processes, the latter being vesicle fusion to plasma membrane, when their contents is effectively expelled. Numerous actors are working to ensure the smooth functioning of exocytosis, most notably proteins, which act and interact at every level of the process in a very strict order.

Proteins are large molecules made of amino acids playing a key role in almost every cell functions. Modeling proteins dynamics and interactions is thus a major component of the cell biology understanding. In particular, proteins are at the heart of every stage of the process, encompassing the vesicle formation, its transport, tethering, docking and, ultimately, fusion.

In this thesis, we study the behavior of three of them: Rab11, Transferrin receptor (TfR) and Langerin. Rab11 is known to be involved in numerous aspects of exocytosis, but pieces of the puzzle are missing, especially regarding the vesicle fusion, and diffusion of the protein. TfR dynamics are well established. As a transmembrane protein, it is constrained to stay in the membrane, thus reducing its degrees of freedom, which facilitates its study. However, while Langerin is also a transmembrane protein, a different behavior is observed in the image sequences, which raises questions about the underlying mechanics. To bring to light quantitative differences between TfR and Langerin, dynamical models must be proposed, estimated and compared. Characterizing the dynamics and discovering the roles and interactions of Langerin or Rab11 is a priority of UMR 144.

1.1.3 Group motion is more than a collection of individual motions

In biological images with smaller magnification than in Part II, several cells move collectively as a single population. As an example, during embryo development, a group of cells divide, move together and interact. Also, in image sequences depicting cell tissues, cells move and deform, thus imposing constraints to their neighbors. In these cases and many others, the motion of single cells is embedded in a higher-level coherent group motion, that of the whole embryo or tissue.

In biology, group motion analysis is generally tackled by estimating the dynamics of each cell individually. Actually, many methods were proposed to study the dynamics of embryos or tissues by detecting and tracking individually each cell. The tracking of individual cells is then performed in order to further derive the relevant information at the group level. However, it is not straightforward to take into account the interactions in individual dynamical models without globally estimating the whole group dynamics. Therefore, it is natural to directly study the dynamics of the whole group, that is, not to differentiate the individual cells.

While not used in biology so far, the latter approach is very common to analyze pedestrian group motions or vehicle traffic, among others. For this reason, we extend the scope of the last part of the thesis to crowd motion analysis, a very active field from where approaches can be adapted to analyze the motion of cell groups.

1.2 Objectives and contributions

1.2.1 Detection of individual spatiotemporal events

As stated earlier, we target two types of dynamics, namely that of independent individuals and that of groups. We first investigate the modeling of a cell regulation process named exocytosis and related control parameters estimation from microscopy image sequences. We specifically focus on the last stage of the process: diffusion. This investigation encompasses the detection of the exocytosis in the image sequence, which itself relies on spot detection because vesicles appear in the images as small spots. To this end, we design a spot detection method with very few parameters which takes advantage of a specificity of the image used for exocytosis characterization, that is, spots to detect have similar sizes in TIRFM image sequences. As opposed to classical spot detectors, we propose an approach for automatically selecting the best detection scale. Then, the detection amounts to a local thresholding which automatically adapts to the local image statistics. The detection sensitivity only depends on a user-specified probability of false alarm, which is set for the whole sequence (or set of sequences) to be processed. For the thresholding to be efficient, the spots in the input image have to be enhanced with a filter at a particular scale, whose choice is critical and is explicitly handled as aforementioned.

1.2.2 Modeling and estimation of fusion dynamics

As for the vesicle fusion characterization, we propose new diffusion models suited to Rab11, TfR and Langerin, which are more realistic than the classically used point source model. Indeed, in the latter, it is assumed that the vesicle size is smaller than the pixel pitch, and that its content is instantaneously released. We relax those hy-

potheses, enabling a spatial extent for the vesicle, and a continuous release, thus better fitting the data. Moreover, in the case of Rab11, it is not known if diffusion occurs in two and/or three dimensions. Therefore, we investigate both possibilities. Once detected, the fusion events with the aforementioned spot detector, the model parameters must be estimated. We rely on an estimation method able to cope with the complexity of the data. Actually, the number of data points is quite low with respect to the noise level, so classical diffusion coefficient estimation methods behave poorly for vesicle fusion dynamics. Furthermore, the biochemical parameters are then compared for different experimental conditions. Parameter statistics are extracted from the whole set of detected fusion events in sequence collections.

This leads us to the last part of the thesis. In most biological applications, the behavior of a single sample (e.g., the dynamics of *one* fusion event in the cell) is not the critical issue. Instead, it has to be compared to the mean, normal or dominant behavior (e.g., the average diffusion coefficient estimated on a set of sequences). Sometimes, the behavior is expected to be similar for all samples, so that coherent or consistent dynamics can be expected. This knowledge can be exploited to improve modeling and estimation, as done to select the optimal scale for spot detection. Therefore, the group properties (velocity statistics, common behavior, count...) are central in the study under concern.

1.2.3 Characterization of group dynamics

To capitalize on the strength of crowd motion analysis methods, we propose to adopt the philosophy in biological image sequence analysis. Specifically, we aim at characterizing the cell crowd dynamics without distinguishing individual cells. However, we do not simply transpose crowd motion analysis methods to biology: to study human crowds or road traffic, existing methods generally use hundreds of frames, or even the whole image sequence, which comes with a very high computation cost. In contrast, we develop instantaneous methods for group dynamics analysis in the remaining of the thesis, that is, methods using only a couple of consecutive frames to characterize the crowd dynamics.

Moreover, in order to gather more information on motion than classical methods based on optical flow, we design a method which provides quantitative and qualitative motion information at the same time. The motion is estimated as a *labeled affine flow*, which is a map conveying both a velocity vector and motion type information. Statistical arguments are used to select a relevant affine motion model at each point of the image, which can be classified owing to simple rules on the affine parameters. Then, the estimates are used to characterize the group motion, or to detect and segment local anomalous dynamics inside the group. To validate the methods, we rely

on available benchmarks, which generally gather crowd and vehicle traffic videos, and on biological images showing groups of cells.

1.3 Organization of the thesis

After an introduction to basic knowledge on the cell and microscopy, the thesis is organized in three parts. Spot detection is handled in Part I, where we focus on fluorescence microscopy images. Then, proteins behaviors after vesicle fusion are modeled and estimated in Part II. Group motions are analyzed in Part III, spanning from microscopy to crowd analysis in videos.

In Part I, an overview of existing spot detection methods is given in Chapter 3. The proposed two-step method is presented in the two following chapters: the scale selection is described in Chapter 4, and the spot segmentation in Chapter 5.

Part II splits into three chapters. A presentation of the exocytosis process is given in Chapter 6, where we provide an overview of existing diffusion models in biology, and present the proposed fusion models. The exocytosis detection is presented in Chapter 7, and the issue of estimating biological parameters is tackled in Chapter 8.

As for Part III, after introducing group motion analysis and related methods, the proposed labeled affine flow is presented in Chapter 9. and motion descriptors for the group motion characterization are introduced. Anomaly detection is addressed in Chapter 10.

Thesis contributions are summarized and possible future works are proposed in Chapter 11. In particular, limits of the acquisition setups and methods designed in this thesis are pointed out, and improvements are proposed.

The thesis organization is synthetically presented below.

Chapter 2. On the cell and its observation 13

The cell organization and components are described to introduce the thesis. The observation model is formulated while we describe the microscopy technique used throughout the thesis, namely TIRFM.

Part I. Spot detection

Chapter 3. The landscape of spot detection 29

Before analyzing dynamics, objects of interest and space-time events must be detected. The (very active) domain of spot detection is introduced and existing spot detection methods are presented. The scale parameters of the existing detectors are outlined.

Chapter 4. Automatic selection of the detection scale 37

We propose a method for automatically selecting the optimal detection scale, which is related to the spot size. Efficiency of the scale selection is shown, and an extension to multiscale detection is proposed.

Chapter 5. Spot segmentation 47

The spots are segmented by thresholding an enhanced image. A local adaptive method is proposed to cope with complex backgrounds. It is thoroughly compared to state-of-the-art spot detectors.

Part II. Individual dynamics**Chapter 6. Modeling of the vesicle fusion dynamics 71**

The biological context of Part II is introduced. We describe the exocytosis process with a focus on vesicle fusion and protein diffusion. Classical diffusion models in biology are presented and discussed. The fusion dynamics of different proteins, namely Rab11, TfR and Langerin, are mathematically modeled.

Chapter 7. Detection of diffusion dynamics 95

In order to further investigate biological parameters of the vesicle fusion, the latter must be detected. We propose detection methods for the three proteins of interest, which exploit the methodology presented in Part I.

Chapter 8. Estimation of the biological parameters 109

After presenting existing methods for estimating diffusion in microscopy images, we propose an estimation method for the proposed vesicle fusion models. Quantitative evaluations and comparisons demonstrate the efficiency and robustness of the method. Real TIRFM images are analyzed to exhibit the differences between TfR and Langerin dynamics.

Part III. Group dynamics**Chapter 9. Labeled affine flow for group motion characterization 127**

Labeled affine flow is proposed as an augmented motion data for group motion analysis. At each point, the motion measurement conveys two types of information: a quantitative motion vector and a qualitative motion type. As a proof of concept, we rely on the proposed labeled affine flow to characterize the group motion, and to recover the dominant paths followed by the group.

Chapter 10. Anomaly detection and localization 147

The proposed labeled affine flow is also used to detect local abnormal events. A patch-based approach allows us to detect and locate anomalies, by comparing local motion class histograms in a center-surround setting.

Chapter 11. General conclusion**163**

Main results are summarized. We discuss the contributions of the thesis with a critical analysis of all the developed methods. This leads us to envisage improvements and foresee future work.

2 On the cell and its observation

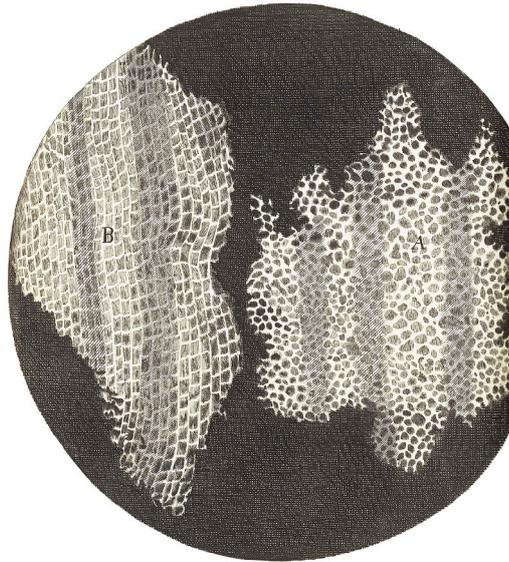
TO ANALYZE image sequences showing cellular dynamics, a basic knowledge of the cell and imaging system is required. We shall not plunge into an exhaustive enumeration of the cell components, but some primary structures must be presented to understand the context of our work, as well as the microscopy technique, which actually relies on genetic processes.

Indeed, numerous biological processes and associated cell compartments take care of parsing genetic information, contained in the cell, to infer which proteins, vital compounds of the cell, should be synthesized. Proteins are precisely at the heart of both the project and observation system.

Along with nucleic acids, lipids and sugars, proteins are the very basic organic components of the cell. They are involved in a wide variety of intra- and inter-cellular processes, where they carry out numerous functions: they generate mechanical forces, transport chemical materials, are key actors of homeostasis and cell defense, catalyze chemical reactions... Regarding the research project, three proteins are studied in particular, namely Transferrin receptor (TfR), Langerin and Rab11. More precisely, we will focus on their dynamics in the so-called exocytosis-recycling process, which is part of the homeostasis regulation process.

To be observed in total internal reflection fluorescence microscopy (TIRFM), proteins we are interested in are genetically associated to fluorescent tags. When excited, they emit light detected by the microscopy setup, thus allowing us to locate and track the proteins.

The chapter is composed of three sections. A quick overview of the cell biology chronology is given in Section 2.1, thus briefly introducing vital processes of the cell. Biological processes involved in protein synthesis and in exocytosis are then detailed in Section 2.2. Finally, the process of fluorescent tagging and TIRFM are presented in Section 2.3.



© Engraving by Robert Hooke [Hooke, 1665]

Figure 2.1 – Sections of a cork bark showing cells

2.1 A bit of history

Today's cell biology knowledge is the result of a 350-year research. Back in the XVIIth century, Robert Hooke – an architect, philosopher, mathematician, physicist, and an optical engineer in his spare time – builds one of the first microscopes ever made to study the porosity of various materials [Magner, 2002]. Analyzing cork bark among other surfaces, he observes a tessellation made of tiles he terms *cells* [Hooke, 1665] by analogy to honeycomb cells (Figure 2.1). A few years later, specifically studying living beings, Antonie van Leeuwenhoek observes cells in various organisms: bacteria, algae, muscle tissues... [Hooke, 1800] But no generalization is made yet.

It will be two centuries before Matthias Jakob Schleiden draws the basement of modern cell theory, defining every plant as a set of cells [Schleiden, 1838]. The following year, Theodor Schwann shows that the theory also applies to animals [Schwann, 1839]. Cell theory basements are completed in 1852 by Robert Remak, who demonstrates that every cell originates from the division of a preexisting cell [Remak, 1852]. In turn, not only is cell a component of every living being, but also a living being itself: cell autonomously carries out the functions necessary to live and reproduce. Among others, it contains the genetic information, synthesizes complex vital molecules from basic compounds, and regulates its content to stay in homeostasis, that is the equilibrium state of its content: sugars, lipids, proteins, ions...

In the second half of the century, genetic sciences emerges, pioneered by Gregor Mendel, who formalizes discrete traits inheritance in 1865 by identifying dominant and recessive traits [Mendel, 1866], and later Friedrich Miescher, who discovers deoxyribonucleic acid (DNA) in 1869 [Miescher, 1871]. The link between both, however,

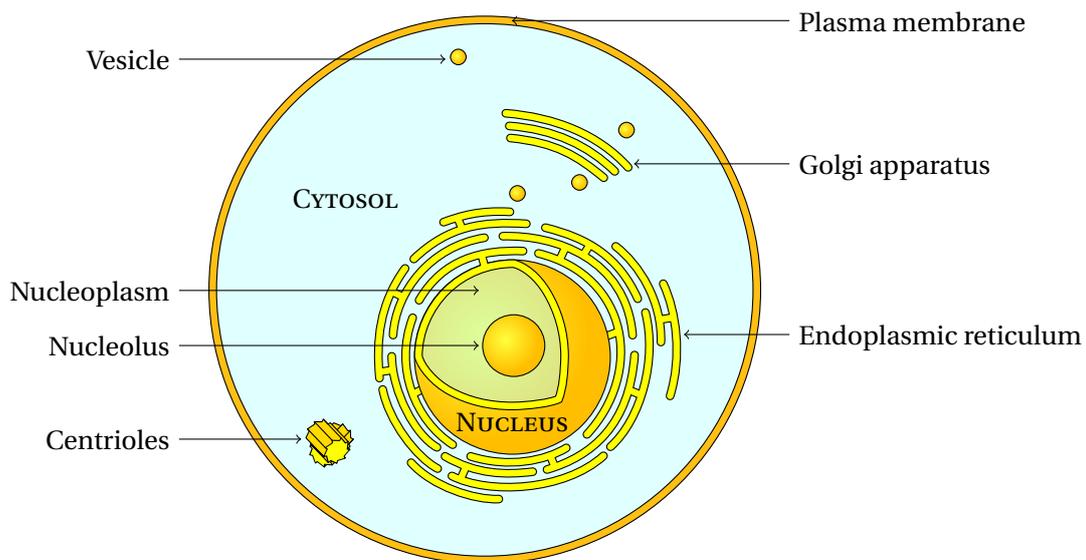


Figure 2.2 – Some components of the cell

was made in the XXth century. In 1902, Theodor Boveri and Walter Sutton advance that the genetic information is contained in the chromosomes. The proof is given in 1915 by Thomas Hunt Morgan [Morgan et al., 1915], a discovery for which he is awarded the Nobel Prize in Physiology or Medicine. In 1944, Oswald Avery finally locates the genetic information in DNA [Avery et al., 1944].

Genetics understanding eventually leads to fluorescence microscopy, which will be described in more details hereafter. In 2008, Osamu Shimomura, Martin Chalfie and Roger Tsien receive the Nobel Prize in Chemistry for the discovery of the green fluorescent protein (GFP) and its application to cell biology [Chalfie et al., 1994]. True to its name, the GFP is a protein which emits green light when excited, e.g., by an incident laser beam. By synthesizing a DNA fragment, they managed to bind GFP to another protein naturally present in the cell, and follow the latter in the living cell using an optical microscope.

The number of discoveries owing to fluorescence microscopy is phenomenal and continues growing fast. Ground-breaking technologies now allow combining the dynamical acquisition of light microscope with the resolution advantage of electron microscopy by freezing the cell during the experiment. In the meantime, the envelope of optical microscopy is expanded every year using multiple-exposure techniques like structured illumination microscopy (SIM) [Neil et al., 1997, Barlow and Guerin, 2007], which shall approach electron microscopy resolution in the near future, while still enabling dynamical recording.

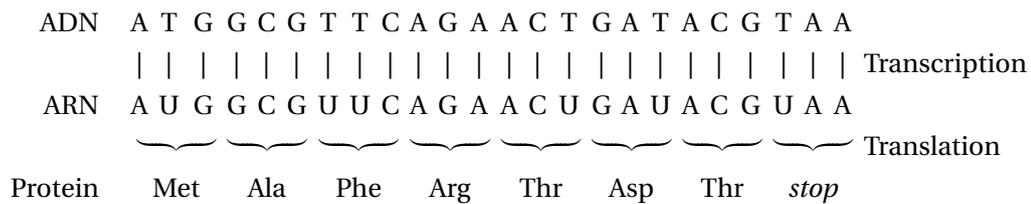


Figure 2.3 – Protein biosynthesis example

2.2 Organization and processes of the cell

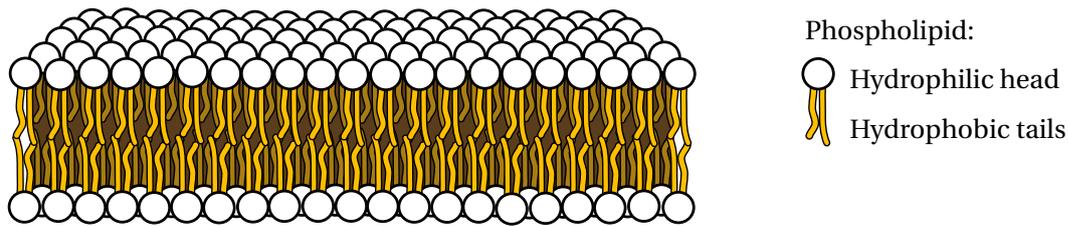
Let us now briefly describe the cell and its organization, by focusing on the relevant biological processes for the thesis. In Figure 2.2, the cell is represented along with some organelles which will be presented below. Cells can be either prokaryotic or eukaryotic. Eukaryotic cells (from Greek *eu*, true, and *karyon*, nucleus), which we study, contain a nucleus and cytoplasm enclosed within a plasma membrane. In contrast, prokaryotic cells (from *pro*, before) lack nucleus. The cytoplasm itself is mainly composed of an aqueous solution named cytosol, in which numerous organelles carry out specific functions.

Additionally, the cell is supported by the cytoskeleton, a set of molecular structures which give its shape and rigidity to the cell. Cytoskeleton is also an actor of numerous dynamical process, some of which will be presented in Section 2.2.3.

2.2.1 From nucleus to cytosol: From genetic information to proteins

Nucleus is the host of most of the genetic material of the cell. A nuclear envelope surrounds the nucleolus and nucleoplasm, the latter containing DNA. Essentially, DNA is a long string of nucleotides, the atomic bricks of genetic information, which can be of four types: Adenine (A), Thymine (T), Cytosine (C) or Guanine (G). Fragments of DNA, genes are sequences of nucleotides that define which proteins are synthesized by the cell, as explained hereafter.

The predominant component of the cell in volume, cytosol is the fluid in which nucleus, organelles and cytoskeleton lie, and the place where proteins are synthesized. The so-called protein biosynthesis process is illustrated in Figure 2.3. To produce a protein, DNA is first transcribed to ribonucleic acid (RNA), a molecular structure similar to DNA that can leave the nucleus to convey genetic information to the cytosol. In RNA, Thymine is replaced with Uracil (U), as shown in Figure 2.3. RNA is then translated by ribosomes, either bound to the endoplasmic reticulum (Figure 2.2) or free in the cytosol. Ribosomes read the genetic information contained in RNA and associate corresponding amino acids which will constitute the protein. Specifically, each amino acid is specified by a group of three RNA bases called codon; translation ends when ribosomes reach a specific *stop* condon.



© Adapted from an illustration by Mariana Ruiz Villarreal, Wikimedia Commons

Figure 2.4 – Plasma membrane lipid bilayer

As a consequence, altering a gene, that is, modifying DNA, can yield a change in the set of proteins present in the cell. It is taken advantage of this phenomenon to tag proteins, in order to observe them in fluorescence microscopy, as explained in Section 2.3.1.

2.2.2 The plasma membrane, a place of exchange

Far from being a simple frontier, the plasma membrane is involved in different aspects of the cell life, from mechanical structure to motility and through chemical regulation, which we are mostly interested in.

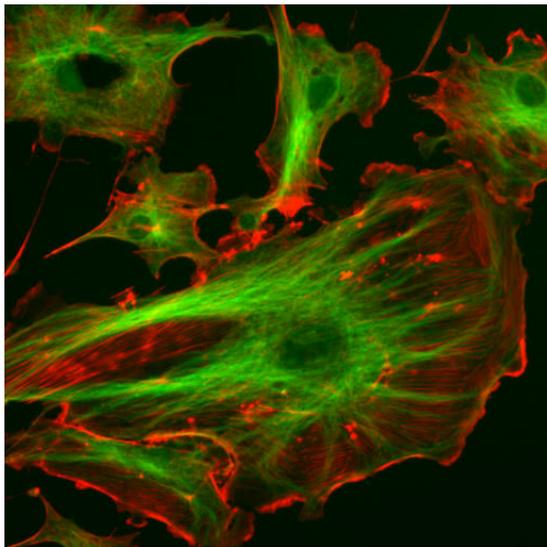
As schematized in Figure 2.4, the membrane is composed of phospholipids, molecules with a hydrophilic-lipophobic head and hydrophobic-lipophilic tails. Phospholipids are arranged in two layers where the hydrophilic head points toward the aqueous cytosol or in the aqueous medium outside the cell, which isolates the hydrophobic tails inside the membrane.

Every chemical flow entering or leaving the cell must pass through the plasma membrane. Thus, the membrane is of first importance in the protection from extra-cellular attack and regulation of the cell homeostasis. The plasma membrane is said to be selectively permeable, because only specific material can pass through, sometimes in limited quantity. While small molecules can move across the membrane, either by passing between phospholipids, or by specific biological gates, a more complex process is required to absorb or secrete bigger structures. The latter process, studied in Part II, relies on transport intermediates named vesicles. They will be presented in more details in the next subsection, and later in Part II.

2.2.3 The cytoskeleton and intracellular transport

Apart from the plasma membrane, the cell mechanical properties are due to the cytoskeleton, mainly composed of microtubules and Actin filaments, depicted in Figure 2.5, and of intermediate filaments.

Microtubules are long tubular structures, essentially made of proteins named Tubulin, running from a so-called microtubule organization center extending around



Bovine pulmonary artery endothelial cells:

- Microtubules
(Tubulin stained with Bodipy FL goat anti-mouse IgG)
- Actin filaments
(F-Actin stained with Texas Red X-Phalloidin)

Intermediate filaments are not visible.

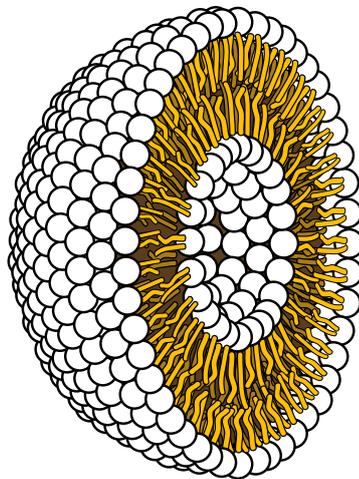
© ImageJ sample image

Figure 2.5 – The cell cytoskeleton

centrioles (Figure 2.2) to the peripheral layers of the cell, where they meet Actin through intermediate filaments. As for Actin filaments, they are mainly observed at the plasma membrane, where they form a dense meshwork, and are involved in various cellular processes.

Indeed, functions of the cytoskeleton are not restricted to structural properties. For example, dynamics of the cytoskeleton are involved in the cell motility and cell division in the mitosis process. However, we are particularly interested in the conveying functions of the microtubules and Actin filaments, and more precisely in vesicle-mediated exocytosis.

In the second part of the thesis, we study the dynamics of vesicles during exocytosis. A vesicle is a closed bag formed by a lipid bilayer membrane similar to the plasma membrane, as depicted in Figure 2.6. Cargo molecules are the ones which are trans-



© Adapted from an illustration
by Mariana Ruiz Villarreal,
Wikimedia Commons

Figure 2.6 – Vesicle section showing the lipid bilayer

ported by the vesicle, from a donor compartment, or from the extracellular space in the case of endocytosis, to a target compartment, or to the extracellular space in the case of exocytosis. To this end, the vesicle forms from the membrane of donor organelle at the beginning of the transport, and fuses to the target membrane at the end of the transport. In between, the vesicles is moved by molecular motors along microtubules, which act as highways in the cell. Then, in the case of exocytosis transport, in order to reach the plasma membrane for expelling its content, the vesicle has to make its way through the Actin meshwork, before fusing to the membrane to release its content.

Furthermore, several types of structures travel with the vesicle, including:

- Transmembrane proteins, which span across the vesicle membrane during transport and are released in the target membrane after fusion;
- Vesicular motors, which pull the vesicle along the cytoskeleton.

Vesicular transport will be described in more details in Chapter 6, as it is a fundamental component of exocytosis, that is the subject of Part II. Specifically, we will focus on two transmembrane proteins, TfR and Langerin, and on Rab11, a protein which is associated to the vesicular motor during transport.

2.3 TIRFM for imaging membrane dynamics

2.3.1 Fluorescence microscopy

As the name implies, total internal reflection fluorescence microscopy (TIRFM) is part of the fluorescence microscopy acquisition techniques. The latter is a type of microscopy where the structure of interest, some protein in the frame of the thesis, is tagged with a fluorescent compound called fluorophore. By tagged, we mean that the fluorophore, here fluorescent proteins, bounds to the protein of interest, so that locating the fluorophore roughly corresponds to locating the protein itself. As mentioned in Section 2.1, fluorophores emit photons when excited by incident light. Thus, in fluorescence microscopy, the tag itself emits the light which is collected by the microscope sensor. Structures of interest can hence be selectively observed, while all other structures remain invisible to the microscope sensor.

As for the imaging device, the principle of the fluorescent microscope is schematized in Figure 2.7, which represents an inverted microscope, meaning that the objective points upward. The objective top is immersed in some optical oil that has the same refractive index n as the front lens element and cover slip, about 1.5. This ensures that the contact surface between the cell and cover slip is the sole optical interface between the microscope objective and the cell itself.

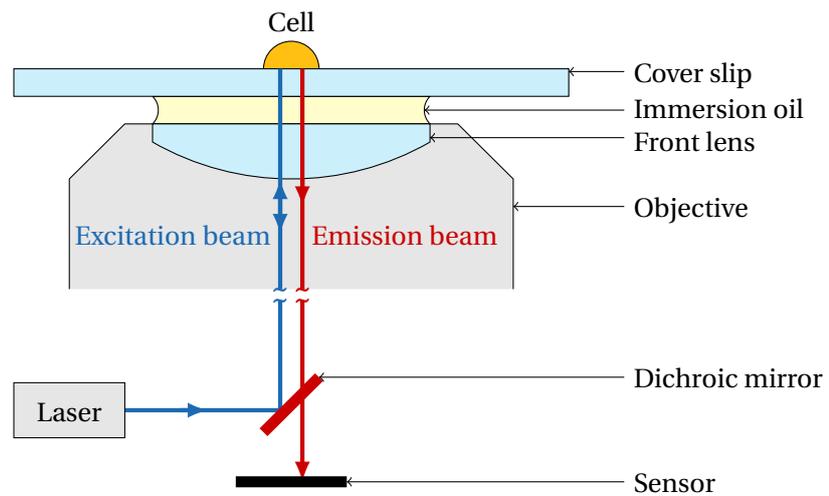


Figure 2.7 – Principle of fluorescence microscopy

In order to illuminate the fluorophores, an excitation laser beam (or other sources of illumination), represented in blue in Figure 2.7, passes through the microscope objective and reaches the cell. Fluorophores which are illuminated by the beam get excited and then emit photons. A portion of the light emitted by the fluorophores, represented in red, points toward the microscope sensor. A fundamental property of the fluorophore is that the emitted beam has a longer wavelength than the excitation beam. Thanks to this, a dichroic mirror allows separating the excitation and emission wavelengths, in order to prevent the excitation beam to hit the sensor after some reflection. In the end, only the photons emitted by the fluorophore are acquired by the sensor, corresponding to tagged structures.

Nevertheless, fluorescence has drawbacks as well, principally phototoxicity and photobleaching, which are closely related to the laser beam power. Phototoxicity is the killing of certain cellular structures due to the relatively high-energy photons hitting them, which affects cell functioning and eventually leads to its death. In order to preserve the functions of the cell during the image sequence acquisition, a compromise must be made between the signal-to-noise ratio (SNR) and toxicity, that is, the laser power must be carefully set.

Photobleaching is the phenomenon by which excited fluorophores can either emit a photon or enter a chemical reaction which eventually kills its fluorescence ability. Over time, more and more fluorophores are bleached and do not emit light anymore. Therefore, bleaching is visible in the images as a slow fading of global intensity. Hopefully, we will see that it is mostly negligible in TIRFM.

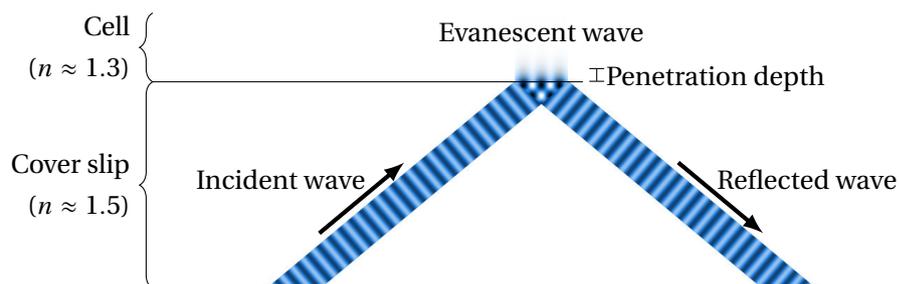


Figure 2.8 – Principle of total internal reflection generating the evanescent wave

2.3.2 Total internal reflection

In TIRFM, the fluorophores are excited by a so-called evanescent wave [Axelrod, 2008]. The latter forms inside the cell when the incident light beam is totally reflected at the interface between the cover slip and the cell itself. Total reflection occurs when the incident angle of the laser beam exceeds a critical angle. This is illustrated in Figure 2.8. Energy of the evanescent wave decreases exponentially with the distance to the cover slip, so that practically no energy is transmitted above a certain distance called penetration depth; its order of magnitude is 10^2 nm [Steyer and Almers, 2001].

Therefore, in TIRFM, only fluorophores close to the coverslip are illuminated and excited. This allows to make visible only the molecules that are inside or very close to the plasma membrane – approximately up to the penetration depth. As a consequence, TIRFM is particularly well suited to study the dynamics occurring at the plasma membrane, such as vesicle fusion studied in Part II [Reichert and Truskey, 1990, Deng et al., 2009, Letinic et al., 2010, Burchfield et al., 2010].

Yet, TIRFM scores a better spatial and temporal resolution than classical optical microscopy techniques. Moreover, as the light energy transmitted to the cell is confined to a very thin region, the power of the incident laser beam can be quite low to get a sufficient intensity at the plasma membrane. Compared to classical fluorescence microscopy, the energy transmitted to the plasma membrane can be higher without being toxic, so TIRFM features a very high SNR among fluorescence microscopy techniques. For the same reason, photobleaching is much reduced, so that when using local approaches which rely on few consecutive frames, it can merely be neglected.

To acquire a color image (e.g., Figure 2.5), fluorophores with different emission wavelengths can be used to tag different proteins. Using a color splitting technique named Dual view [Gidon et al., 2012], the wavelengths are then projected side-by-side on the sensor using a set of mirrors or prisms. This technique can be employed to compare the localization of different structures, or to use one structure property to ease working on another structure, as done in Chapter 7.

2.3.3 Point spread function

Point spread function (PSF) is the optical transfer function. Even for an ideal microscope, the image of a point in focus by the optical system is not a point but rather a blurred spot. The more out-of-focus the point, the wider the spot. Theoretically, a point in focus is projected through a pinhole or diaphragm as an Airy disc [Airy, 1835, Marian et al., 2007]. Naturally, the model becomes more complex with more complex optical formulas.

Moreover, even with a theoretical, ideal lens, a PSF model should take into account the discrete nature of the microscope sensor. Indeed, the photons collected at a given pixel are spread over the photosite area. While small in absolute terms, the photosite area is not negligible compared to the PSF width. Therefore, instead of considering a sampled Airy function to model the amount of light gathered by a photosite, the function should be spatially integrated over the pixel surface [Small and Stahlheber, 2014].

Then come optical aberrations introduced by the lens, such as astigmatism and coma. Not only are these aberrations hard to infer from optical specifications, but they also vary spatially [Aguet, 2009]. The variation theoretically only depend on the distance from the optical center, but real optics are not precisely enough manufactured to stick to this assumption. Therefore, in order to precisely model the PSF over the whole image, classical methods consist in fitting parametric models over the image domain, and then interpolating the parameter field in order to get a continuous PSF map [Aguet, 2009].

In practice, apart from very specific applications as deconvolution [Sibarita, 2005, Sarder and Nehorai, 2006] or super-resolution [Carrington et al., 1995, Zhang et al., 2006], a Gaussian approximation is often adequate to model the PSF in TIRFM images [Small and Stahlheber, 2014], so we rely on this approximation throughout the thesis.

Furthermore, in TIRFM, objects at positive heights are out of reach for the evanescent wave. Therefore, they do not appear in the image and all visible objects are in focus; if not, they are at about the same distance from the focus plane. Therefore, objects of similar size in the cell are projected as spots of similar size in the TIRFM image.

2.3.4 Digital image model

Notation and units

The image sequence, denoted I , is a succession of digital photographs termed frames. Frame at time t will be denoted as $I(t)$. When not otherwise stated, temporal unit will

Table 2.1 – Biological and microscope orders of magnitude

Description	Dimension	Weight*	Duration
Cell (diameter)	$\approx 10^2 \mu\text{m}$	$\approx 1 \text{ ng}$	
Nucleus (diameter)	$\approx 10^1 \mu\text{m}$		
Vesicle (diameter)	$\approx 10^2 \text{ nm}$		
Actin filament (diameter)	7 to 8 nm		
Microtubule (diameter)	15 to 25 nm		
Membrane (thickness)	4 to 5 nm		
Membrane phospholipid		< 1 kDa	
Rab11		24 kDa	
Langerin		37 kDa	
TfR		85 kDa	
mCherry		29 kDa	
pHluorin		27 kDa	
Spatial optical resolution	200 nm		
Physical pixel side	16 μm		
Pixel side in the image	160 nm		
Penetration depth	$\approx 10^2 \text{ nm}$		
Exposure time			100 ms
Frame rate			10 f/s
Sequence length			30 to 120 s

$$1 \text{ kDa} \approx 1.66 \times 10^{-21} \text{ g}$$

*Source: PhosphoSitePlus (www.phosphosite.org)

be the frame period (f), or frame in short, in the remaining of the thesis, such that $I(0)$ is the first frame of the sequence and $I(t + 1)$ is the frame following that of time t .

The photosites collect photons and convert them into electrons. Photosites are arranged in a grid, and are represented by their Cartesian coordinates x and y . The distance between horizontally and vertically adjacent pixels is the pixel pitch, denoted px , which is the natural distance unit in the image. This way, pixels have integer coordinates. The set of pixel locations, called domain, is denoted Ω_I and is therefore a subset of \mathbb{Z}^2 .

The intensity of a pixel $p = (x, y) \in \Omega_I$ in frame $I(t)$ is denoted $I(p, t)$. It is a function of the number of photons $n(p, t)$ received by the associated photosite from time t to time $t + \Delta t$, where Δt is the exposure time. In the images we deal with, there is a negligible time interval between consecutive exposures, so Δt is approximately the frame period: $\Delta t \approx 1 \text{ frame}$.

Mixed Poisson-Gaussian process noise model

A mixed Poisson-Gaussian (MPG) model is often proposed to account for noise in fluorescence microscopy images [Zhang et al., 2007]. Actually, this model is not restricted to fluorescence microscopy, and is generally assumed in digital images when the following two random components are taken into account:

- Photon noise, which results from the stochastic process of photon emission and conversion;
- Electronic noise, which is generated by the sensor and electronics.

In addition to those components, quantification noise owes to the storage of the intensity signal as a digital image where each pixel only takes discrete values. As we are dealing with image sequences scoring high bit depth (the intensity is quantified over 2^{16} levels), this component is negligible.

In fluorescence microscopy, photons are emitted when an excited fluorophore goes back to its fundamental state. The time before photon emission, or lifetime, is a stochastic process, in which the remaining time before emission does not depend on the elapsed time from excitation. However, when a group of fluorophores are excited, the number of excited atoms at a certain time is expected to decay exponentially. As a result, the number $n(p, t)$ of photons emitted toward pixel p during time interval $[t, t + \Delta t]$ obeys a Poisson statistic. This accounts for the Poisson component of the noise model, which is hence signal-dependent.

As for electronic noise, it results from different sources, yielding the following main components:

- Dark-current noise reflects the generation of electrons by the sensor even in the absence of photons, due to the non-null electron excitation in the sensor and electronics;
- Similar to photon noise, electronic shot noise accounts for the quantization of the electric potential levels, as the electric charge representing the pixel signal is quantified as the number of electrons;
- Flicker noise, which depends on the signal frequency, arises from direct current passing through the photosite.

Generally, as it is extremely difficult to tell apart the different intermingled components, the whole electronic noise is modeled as a Gaussian process [Boulanger, 2007].

Finally, the signal is amplified by a multiplicative gain g_0 . As a result, the intensity $I(p, t)$ of a pixel p at time t is modeled as [Boulanger, 2007]:

$$I(p, t) = g_0 n(p, t) + \varepsilon(p, t), \quad (2.1)$$

where g_0 is the system gain, $n(p, t)$ is the number of photons received at pixel p between times t and $t + \Delta t$, which follows a Poisson law, and $\varepsilon(p, t)$ is the sample of a Gaussian distribution of spatially constant mean μ_ε and variance σ_ε^2 .

Variance stabilization

Most image processing methods assume a white Gaussian noise. Therefore, they are not well suited to the MPG model. To deal with fluorescence microscopy images, an intensity transform is often used, which gives a white Gaussian noise in the transformed images [Zhang et al., 2007].

In the remaining of the thesis, we will rely on the generalized Anscombe transform (GAT) \mathcal{T} , defined as [Boulanger et al., 2010b]:

$$\forall p \in \Omega_t, \forall t \in \mathbb{N}, \mathcal{T}(I(p, t)) = \frac{2}{g_0} \sqrt{g_0 I(p, t) + \frac{3}{8} g_0 + \sigma_\varepsilon^2 - g_0 \mu_\varepsilon}. \quad (2.2)$$

Parameters g_0 , μ_ε and σ_ε^2 will be estimated with a patch-based method proposed in [Boulanger et al., 2010a].

Other variance stabilization methods exist, as the multiscale variance stabilizing transform (MS-VST) proposed in [Zhang et al., 2007], but the GAT is computationally lighter, and is sufficient in the situations we encounter in the thesis, as demonstrated by the experimental results reported in the remaining of the thesis.

I Spot detection

3

The landscape of spot detection

SINCE THE EARLY TIME of protein tagging with green fluorescent protein (GFP) [Chalfie et al., 1994], microscopy investigations at the single cell level have been faced with the problem of determining the location and behavior in space and time of spots, such as microtubule end tips, adhesion molecular complexes, or vesicles as illustrated in Figure 3.1. Detecting such subcellular particles in fluorescence microscopy is indeed of central interest for further quantitative analysis as particle counting [Byun et al., 2006], particle pattern recognition [Jackson et al., 2011], particle tracking [Sbalzarini and Koumoutsakos, 2005, Miura, 2005, Jaqaman et al., 2008, Chenouard et al., 2014] or dynamics classification [Sage et al., 2005, Boulanger et al., 2010a, Sironi et al., 2011, Basset et al., 2014b]. All these subcellular analyses

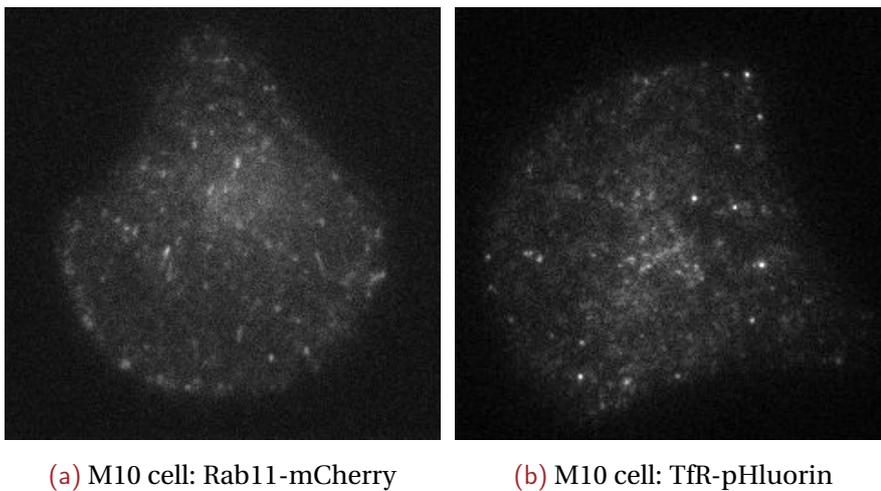


Figure 3.1 – Cell images depicting particles of similar scale. (a,b) Tagged vesicles (bright spots) are of almost constant size over the image. Rab11 is tagged with mCherry in (a), (b) TfR is tagged with pHluorin in (b).

start with a reliable, accurate and efficient detection of particles in fluorescence microscopy images.

Our goal is to segment exocytotic vesicles in cell images acquired in total internal reflection fluorescence microscopy (TIRFM). Among fluorescence microscopy image modalities, TIRFM is the perfect tool to investigate processes occurring close to or at the cell surface such as endocytosis and exocytosis processes [Reichert and Truskey, 1990, Deng et al., 2009, Letinic et al., 2010, Burchfield et al., 2010]. The physical size of exocytotic vesicles spans across a limited range. Given the limited depth of field (DOF) of TIRFM, the variation of the scale of these fluorescently labeled objects in the 2D images is also limited. In this part, we will focus on M10 cells showing the cargo proteins Langerin and Transferrin receptor (TfR) tagged with pHluorin, or the Rab11 GTPase tagged with mCherry. These proteins are associated to transport intermediates such as vesicles recycling to the cell surface and appearing as bright spots, which can be round or elongated, as depicted in Figure 3.1. Another application of the presented method could be the identification, detection and quantification of adhesion molecular complexes, in cells migrating or not. These biological architectures are relatively small and regular at the single cell, composed of multiple molecular partners.

As a consequence, it is worth developing a spot detection method able to automatically find the average object size or the most frequent one. We propose a segmentation framework with automatic scale selection and local adaptive thresholding. Our method exploits the Laplacian of Gaussian (LoG) of the intensity image and automatically detects the characteristic scale of the objects of interest. To cope with inhomogeneous background, thresholding is adapted to local statistics, while a single probability of false alarm (PFA) is set for the whole image or even the collection of images to be processed. In short, we will automatically infer from image data the optimal parameters usually left to the user guidance in other methods, that is, LoG scale and detection threshold. We name adaptive thresholding of LoG images with auto-selected scale (ATLAS) the method described in this part.

ATLAS comprises several significant improvements and extensions compared to the preliminary method SLT-LoG we introduced in [Basset et al., 2014a]:

- We now resort to a discrete filter for the scale-space representation and we can deal with any arbitrary scale, i.e., with scales of any precision;
- We have designed four original scale selection criteria;
- We have produced and made publicly available a new benchmark dataset for spot detection methods;
- We have conducted an extended comparative evaluation with existing methods on several datasets, and we have evaluated our method on a larger range of real images.

While our primary goal is to detect exocytic vesicles in two-dimensional TIRFM images, the ATLAS method can be applied to other types of images as well, provided objects to be detected are of similar size in the image or of a couple of sizes at most.

Comparisons of spot detection methods were reported in [Smal et al., 2010, Ruusuvaari et al., 2010], providing with a broad overview of state-of-the-art methods. However, the dataset used in these previous experiments remains limited in terms of content and challenges. Indeed, real TIRFM images are far more complex than images of this dataset, specifically, the signal-to-noise ratio (SNR) is generally lower in real images and objects to be detected are smaller and often darker. We have then constructed a more realistic and more challenging dataset with ground truth to quantitatively evaluate and compare methods. In addition, we have used images supplied by the simulators designed in [Rezatofghi et al., 2013, Boulanger et al., 2009].

Part I is organized as follow. The present chapter gives an overview of existing spot detection methods in fluorescence microscopy, and compares the different approaches for setting the detection scale(s). Our automatic scale selection is presented in Chapter 4, and the actual adaptive detection is described in Chapter 5, where we also compare our detection results with those of state-of-the-art methods.

Existing spot detectors

In [Smal et al., 2010], the authors provide a broad panorama of spot detection methods, and thoroughly evaluate the performance of a dozen methods. As explained by Smal *et al.* [Smal et al., 2010], the common detection framework consists in first denoising the image and enhancing the spots to be detected. Then, highest or lowest values of the enhanced signal, corresponding to spots, are extracted.

The simplest way of detecting spots in a gray level image is to threshold the image intensities from the intensity histogram. The threshold value can be automatically selected by techniques such as Otsu's method [Otsu, 1979] or entropy minimization [Kapur et al., 1985, Sahoo et al., 1997]. However, a single global threshold cannot tackle complex images where variation in background intensities may exceed spot intensity magnitude. Therefore, numerous space-varying thresholding methods were proposed [Sahoo et al., 1988, Sezgin and Sankur, 2004]. In particular, local threshold values are deduced from local statistics to detect cell nuclei in [Phansalkar et al., 2011].

More advanced methods, such as detectors based on top hat filter (TH) [Bright and Steel, 1987, Breen et al., 1991] or LoG filter as in the spot enhancing filter (SEF) method [Sage et al., 2005], not only smooth the image, but also enhance the underlying signal. More specifically, the LoG filter (which we will rely on) is a band-pass filter which enhances objects of a particular size, reduces noise and lowers low-frequency background structures. Sage *et al.* [Sage et al., 2005] empirically observed that the

LoG filter is close to the optimal whitened matched filter for Gaussian spots in fluorescence microscopy images, that is, the SNR of the filtered image is maximized at the spot center. Yet, the choice of the LoG variance is critical and highly dependent on the spot size. Similarly, the bandwidth of the TH filter is adjusted with two critical parameters, the top and brim radii. Ideally, they should correspond to the spot size and distance between neighboring spots, respectively. In the so-called morphological top hat filter (MTH) version of TH [Soille, 2003], the image background is estimated by an opening operation which removes objects smaller than the structuring element. In order to reduce noise, a Gaussian blur is initially performed. The background estimate is then subtracted to the image to detect spots by thresholding.

In [Olivo-Marin, 2002, Zhang et al., 2007], an iteratively undecimated wavelet transform (IUWT) [Mallat, 1989] of the image is exploited to detect objects of various sizes. A wavelet multiscale product (WMP) operation is performed in [Olivo-Marin, 2002], which consists, for every point, in multiplying the wavelet coefficients of different scales to reveal correlations across the scales. Indeed, from a given wavelet scale, spots respond more strongly to IUWT than uncorrelated noise. However, for low SNRs, noise has a higher response than spots at smallest scales, inducing wrong detections. Hence, smallest scales – up to a characteristic scale – must be discarded to lower the false detection rate. Finally, the WMP map is thresholded to get the binary detection map. The multiscale variance stabilizing transform (MS-VST) method relies on variance stabilization to rule out insignificant coefficients of the IUWT [Zhang et al., 2007]. Then, the image is reconstructed without taking into account the coarsest scale, corresponding to the background structures, nor the smallest ones corresponding to noise. The spots are finally detected by thresholding the reconstructed image. Therefore, with both IUWT-based methods, the set of wavelet scales must be chosen accordingly to the spot size.

Finally, h -dome (HD) methods [Smal et al., 2008, Rezatofghi et al., 2012, Vincent, 1993] detect local maxima, called domes, in a LoG- or Gaussian-filtered image. The kernel must be chosen smaller than the spots. Peaks of the filtered image with an amplitude greater than a given height h (hence, the name of the method) are extracted. The so-built “dome map” comprises small domes corresponding to noise, domes corresponding to spots, and large domes corresponding to background structures. To discard irrelevant large and small domes, samples are generated according to the domes map seen as an importance sampling function. Domes containing too few samples are removed since they probably correspond to noise. Domes where samples are too scattered are also removed, because they probably correspond to large background structures. Thus, the maximum dome size must be carefully set. However, the objects to detect do not often have the same magnitude h , so that the method sometimes merges very bright neighboring spots, and sometimes misses dark spots.

Table 3.1 – Main parameters of state-of-the-art spot detection methods (see main text for acronym meaning)

Method	Main parameters
TH [Bright and Steel, 1987]	<i>Top-hat radius*</i> Distance between spots Intensity threshold
HD [Smal et al., 2008]	<i>Gaussian variance*</i> Maximum dome radius Dome height
WMP [Olivo-Marin, 2002]	<i>Minimum and maximum scales*</i> Intensity threshold
IFD [ter Haar Romeny, 2003]	<i>Gaussian variance*</i> Intensity threshold
MTH [Soille, 2003]	<i>Smoothing scale*</i> <i>Structuring element radius*</i> Distance between spots Intensity threshold
SEF [Sage et al., 2005]	<i>Gaussian variance*</i> LoG threshold
MS-VST [Zhang et al., 2007]	<i>Minimum and maximum scales*</i> False discovery rate Intensity threshold
MPHD [Rezatofighi et al., 2012]	<i>Gaussian variance*</i> Maximum dome radius Gradient norm threshold
AB [Jiang et al., 2007]	Feature set Feature parameters Size threshold
FDA [Smal et al., 2010]	<i>Patch size*</i> Size threshold
LR-MRF [Ruusuvuori et al., 2012]	Feature set <i>Feature scales*</i> Sparsity parameter Regularization parameter

*Equivalent to scale parameter

To tackle this problem, Rezatofighi *et al.* [Rezatofighi et al., 2012] proposed a method called maximum possible height dome (MPHD) for locally detecting the best height threshold h . Then, the norm of the spatial image gradient is thresholded, which is more robust to strong background variations than directly thresholding intensity.

Table 3.1 collects the main parameters of the aforementioned methods to be set by the user. All these methods have in common one or two critical scale parameters

whose optimal value is closely related to the size of the objects to detect. If the detection scale is chosen too small, overdetection occurs due to noise. If it is set too high, objects are smoothed out or merged when close to each other. Fortunately, the scale parameter, whatever it is, can often be inferred from image data. We will address this issue in Chapter 4. Moreover, let us point out that our characteristic scale detection method is not limited to our detection framework. Indeed, it could be applied as a preprocessing step to most of the aforementioned methods involving a scale-related parameter.

Most methods end up delivering a binary detection map after thresholding a filtered or reconstructed image. However, when processing fluorescence microscopy sequences, the statistics of the image may vary in time due to photo-bleaching, so that one threshold should be set for each frame according to the current image intensity range. Obviously, this approach is not applicable to sequences containing hundreds of frames, or to a large datasets containing images of various dynamic range. In contrast, as described in Chapter 5, we propose a locally adapted threshold automatically inferred from local intensity statistics. The user on his/her side only fixes once for all a PFA value which can be used for all the images of the conducted experiment.

4 Automatic selection of the detection scale

THE OBJECT SCALE selection step consists in determining the most represented scale in the input image I , that is the scale that most spots share. To this end, we rely on the framework developed by Lindeberg for scale-space analysis [Lindeberg, 1993]. In the remaining of the thesis, I will denote a 2-dimensional image, that is $I: \Omega_I \subset \mathbb{Z}^2 \rightarrow \mathbb{R}$, where Ω_I is the image domain. In this thesis, we have focused on 2D TIRFM images, but the proposed spot detection method can be straightforwardly extended to three dimensions.

4.1 Scale-space representations

Definition 1. The *scale-space representation* L of an image I is a 3-dimension map defined in [Lindeberg, 1993] as:

$$\begin{aligned} L: \Omega_I \times \mathcal{S} &\rightarrow \mathbb{R} \\ (p, s) &\mapsto (G_s * I)(p), \end{aligned} \quad (4.1)$$

where p is a point in Ω_I , $\{G_s\}_{s \in \mathcal{S}}$ is a family of convolution kernels of *scale* s , and the set \mathcal{S} of scales is a subset of \mathbb{R}_+^* .

Gaussian kernels are often proposed to build the scale-space representation. As a matter of fact, they are the best kernels for representing continuous signals [Lindeberg, 1993]. However, in the case of discrete signals such as digital images, sampled Gaussian kernels are not optimal, because they fulfill the scale-space representation for restrictive conditions on \mathcal{S} only.

Proposition 1 (from [Lindeberg, 1990]). If $\{G_s\}_{s \in \mathcal{S}}$ is a family of sampled Gaussian kernels of variance s , L is a scale-space representation of I if and only if $\mathcal{S} = \{s_0 r^n | n \in \mathbb{N}\}$ where s_0 is a strictly positive real number and r an odd integer.

Thus, the smallest possible ratio r between two consecutive scales is 3 when using sampled Gaussian kernels. We first adopted this approach in a preliminary work described in [Basset et al., 2014a] with $\mathcal{S} = \{1, 3, 9, 27, 81\}$. However, since vesicles are usually under-resolved in TIRFM images, small scales (say, between 1 and 3) are of primary interest. Hence, we must be able to deal with scale sampling finer than in [Basset et al., 2014a] to further improve detection results. As a consequence, we prefer to employ a non-Gaussian kernel, called discrete analogous of the Gaussian kernel or *discrete Gaussian* in short. It will allow us to introduce scales of arbitrary precision. It is a separable filter, based on the family of modified Bessel functions of integer order $\{B_n\}_{n \in \mathbb{N}}$ ¹. When exploiting this filter family, the scale-space representation holds with scales arbitrarily close to each other [Lindeberg, 1990]. Precisely, we will consider the scale set $\mathcal{S} = \{s_0 r^n | n \in \mathbb{N}\}$, where the scale ratio r is a strictly positive real number and $s_0 = 1$. In our TIRFM images, the pixel size (about 160 nm) is smaller than the optical resolution (about 200 nm), so that no objects of scale smaller than 1 can be found, thus the choice of $s_0 = 1$. The one-dimensional discrete Gaussian kernel G_s of scale s is given by:

$$\forall n \in \mathbb{Z}, G_s(n) = e^{-s} B_n(s). \quad (4.2)$$

The two-dimensional kernel is obtained by convolution, resulting in the following scale-space formulation:

$$\forall (p, s) \in \Omega_I \times \mathcal{S}, L(p, s) = G_s * G_s^T * I(p), \quad (4.3)$$

where T denotes the transposition operation. This formula straightforwardly extends to higher dimension.

The Bessel functions have no closed form and B_n is defined as the canonical solution of the following differential equation [Abramowitz and Stegun, 1972]:

$$\forall z \in \mathbb{C}, z^2 \frac{\partial^2 B_n}{\partial z^2}(z) + z \frac{\partial B_n}{\partial z}(z) + (z^2 - n^2) B_n(z) = 0. \quad (4.4)$$

In practice, recursive approximations of the Bessel functions (such as the one proposed in [Abramowitz and Stegun, 1972]) are precise enough to cope with our problem.

The discrete Gaussian filter is a low-pass filter which smooths out a certain amount of noise depending on the filter scale. In order to enhance the objects to detect, we apply the Laplacian operator Δ . For two-dimensional images, we use the most isotropic Laplacian kernel [Lindeberg, 1990] defined as:

$$\Delta = \begin{pmatrix} 1/6 & 2/3 & 1/6 \\ 2/3 & -10/3 & 2/3 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}. \quad (4.5)$$

¹The Bessel functions are historically denoted I_n but we want to avoid confusions with the image sequence I

To save computation time and owing to convolution commutativity, we apply the Laplacian operator once for all to the input image, before applying the scale-space operator. Moreover, the contrast of the LoG map decreases when scale increases. A normalization operation is performed to countervail this effect, which gives the following *normalized discrete LoG filter* H :

$$\begin{aligned} \forall (p, s) \in \Omega_I \times \mathcal{S}, H(p, s) &= \alpha(s) \Delta L(p, s) \\ &= \alpha(s) (G_s * G_s^T * \Delta I)(p), \end{aligned} \quad (4.6)$$

where α is the normalization function. Lindeberg proposed to normalize the discrete LoG kernel with its L_1 -norm [Lindeberg, 1990], which gives:

$$\alpha^{-1}(s) = \sum_{n=-\infty}^{+\infty} |\Delta G_s(n)|. \quad (4.7)$$

Due to the presence of absolute values, the normalization function (4.7) is not derivable at some particular points of the Bessel functions. Instead, we normalize with the L_2 -norm of the discrete Gaussian kernel.

Proposition 2. The proposed normalization function is given by:

$$\alpha^{-1}(s) = \sum_{n=-\infty}^{+\infty} (G_s(n))^2 = e^{-2s} B_0(2s). \quad (4.8)$$

Proof. The discrete Fourier transform (DFT) of G is defined as [Lindeberg, 1990]:

$$\mathcal{F}(G)(\theta) = e^{s(\cos \theta - 1)}.$$

Using Parseval's theorem, we get:

$$\begin{aligned} \alpha^{-1}(s) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(e^{s(\cos \theta - 1)} \right)^2 d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{2s(\cos \theta - 1)} d\theta \\ &= \frac{1}{e^{2s}} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{2s \cos \theta} d\theta. \end{aligned}$$

Then, we conclude using the integral representation of B_0 [Abramowitz and Stegun, 1972]:

$$B_0(2s) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{2s \cos \theta} d\theta.$$

□

We will use this expression in the remainder of the thesis. The scale-space representation is illustrated in Figure 4.1 for a TIRFM image.

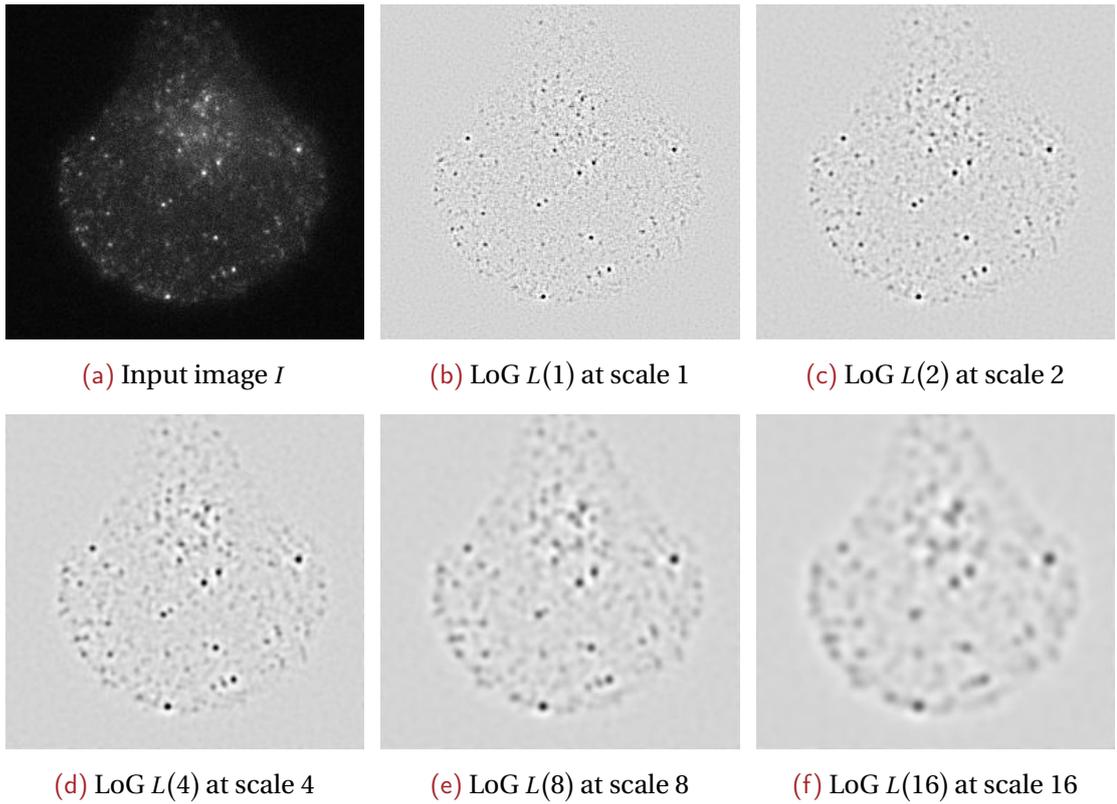


Figure 4.1 – Scale-space LoG-transform L of a real TIRFM image I of M10 cell (Rab11-mCherry)

4.2 Blob distribution and scale selection

Since the LoG filter is negative and spots are bright in the input image, we focus on the negative extreme values of H , which leads us to the concept of (negative) blob as defined in [Lindeberg, 1998].

Definition 2. A blob $b = (p_b, s_b) \in \Omega_I \times S$ is a local minimum of H :

$$\forall (p, s) \in \nu_p, H(p_b, s_b) \leq H(p, s), \quad (4.9)$$

where ν_p is a 3^{N+1} neighborhood of b in $\Omega_I \times S$ for an image of N dimensions.

The set of blobs (resp. blobs at scale s) of an image I is denoted as $\mathcal{B}(I)$ (resp. $\mathcal{B}_s(I)$).

Proposition 3. The set of blobs $\mathcal{B}(I)$ of an image I is unchanged when adding a constant a_0 and/or multiplying I by a positive constant a_1 :

$$\forall a_0 \in \mathbb{R}, \forall a_1 \in \mathbb{R}_+^*, \mathcal{B}(a_1 I + a_0) = \mathcal{B}(I). \quad (4.10)$$

Proof. Let $b = (p_b, s_b) \in \mathcal{B}(I)$ be a blob of I and take $(p, s) \in \nu_p$. Definition 2 gives:

$$H(p_b, s_b) \leq H(p, s).$$

Since $a_1 > 0$, and by linearity of the convolution operation,

$$H_{a_1 I + a_0}(p_b, s_b) \leq H_{a_1 I + a_0}(p, s).$$

Therefore, \mathbf{b} is also a blob of $a_1 I + a_0$, so $\mathcal{B}(I) \subset \mathcal{B}(a_1 I + a_0)$. Analogously, we get $\mathcal{B}(a_1 I + a_0) \subset \mathcal{B}(I)$ and we conclude that $\mathcal{B}(a_1 I + a_0) = \mathcal{B}(I)$. \square

As we showed in [Basset et al., 2014a], the blob detection itself is not sufficient to satisfyingly extract spots from noisy images. Instead, blob detection will be used to select the LoG scale. In order to select the optimal scale s^* , likely to correspond to the average size of the spots of interest, we have to find the scale at which the blob number is the highest, while discarding blobs due to noise. Precisely, we take into account the number of blobs normalized by the image area, that is, the empirical blob density. The empirical blob density in I at scale s is then given by:

$$\forall s \in \mathcal{S}, \rho_I(s) = \frac{|\mathcal{B}_s(I)|}{|\Omega_I|}, \quad (4.11)$$

where $|\Omega_I|$ is the area of I evaluated in square pixels (px^2).

If the image I was noise-free, the maximum of $\rho_I(s)$ would be attained close to the most frequent spot scale. However, due to noise, some detected blobs do not correspond to real spots. As explained in Section 2.3.4, we suppose that the image is corrupted by an additive Gaussian noise (possibly, after noise variance stabilization). Then, I decomposes into a noise-free image I_0 containing only structures (including spots) and an additive Gaussian noise component g :

$$I = I_0 + g. \quad (4.12)$$

Should I exhibit no structure, the density of blobs detected at each scale s would be $\rho_g(s)$. Therefore, we aim at selecting the scale s^* at which the dissimilarity between $\rho_I(s)$ and $\rho_g(s)$ is the largest.

We first propose four possible scale selection criteria while motivating their formulation. Then, we will compare them to retain one criterion.

First, we can take the difference or the ratio of the densities, which leads to the two following selection criteria, respectively:

$$C_D : s^* = \arg \max_{s \in \mathcal{S}} (\rho_I(s) - \rho_g(s)); \quad (4.13)$$

$$C_R : s^* = \arg \max_{s \in \mathcal{S}} \frac{\rho_I(s)}{\rho_g(s)}. \quad (4.14)$$

Since we are dealing with a white Gaussian noise, pixel intensities in the noise image g are assumed to be independent, and all the pixels have the same probability of being blobs at scale s , which precisely means that the number of blobs follows a binomial distribution [Kingman, 1992]. Then, if the number of blobs is high enough, the

binomial distribution leads to a Poisson distribution (which is the limiting case of the binomial distribution [Kingman, 1992]) for the counting variable $|\mathcal{B}_s(g)|$. Therefore, we can also resort to the symmetric Pearson distance [Pearson, 1900, Belongie et al., 2002] and Kullback-Leibler divergence [Hannig and Lee, 2006] to evaluate the dissimilarity between the two densities. Two more selection criteria can then be formulated:

$$C_P : s^* = \arg \max_{s \in \mathcal{S}} \frac{(\rho_I(s) - \rho_g(s))^2}{2(\rho_I(s) + \rho_g(s))}; \quad (4.15)$$

$$C_{KL} : s^* = \arg \max_{s \in \mathcal{S}} \left(\rho_I(s) - \rho_g(s) + \rho_I(s) \log \frac{\rho_g(s)}{\rho_I(s)} \right). \quad (4.16)$$

While exact for $|\mathcal{B}_s(g)|$, the independence assumption is only an approximation for $|\mathcal{B}_s(I)|$, which is precisely the reason why the proposed scale selection criteria work: we are looking for the scale at which the independence hypothesis does not hold, that is, spots lie at this scale.

To evaluate $\rho_g(s)$, we prefer to avoid a time-consuming denoising step, which could also introduce artifacts. Instead, we simulate g by generating an image \widehat{g} containing only white Gaussian noise.

According to Proposition 3, it is unnecessary to estimate the mean and variance of g , so that we merely sample \widehat{g} from the standard normal distribution. Therefore, we can estimate once for all $\rho_g(s) \approx \rho_{\widehat{g}}(s)$ with $\rho_{\widehat{g}}(s) = |\mathcal{B}_s(\widehat{g})|/|\Omega_{\widehat{g}}|$, instead of generating one noise image per input image. To limit quantification noise in the blob counting, it is preferable to generate a large image \widehat{g} . In practice, we use an image of size 4096 x 4096. In our experiments, using an even larger noise image did not impact the scale selection results.

As shown in Figure 4.2b, blob densities in I and \widehat{g} are close to each other except in the vicinity of the spot characteristic scale s^* . Thus, the optimal scale is in general correctly determined by all the criteria. We have studied the sensitivity of the method to the ratio r between consecutive tested scales. Results on a real TIRFM image are presented in Figure 4.2c-f for the four criteria. The sensitivity of the scale ratio parameter r is very low. Nevertheless, using an extremely small ratio value such as 1.1 can give bad results due to the limited global amount of blobs. To achieve a precise enough while reliable scale selection, we fix the scale ratio r to 1.2 in all the experiments and we take $\mathcal{S} = \{1, 1.2, 1.44, \dots, 18.49, 22.19\}$.

To evaluate the scale selection precision of the different criteria, we have generated synthetic images with the ImageJ plugin presented in [Smal et al., 2010]. They contain Gaussian spots of various sizes and are corrupted by a Poisson noise for different SNRs, as illustrated in Figure 4.3. Table 4.1 demonstrates the strong correlation between the size of the spots (defined by the Gaussian variance) and the selected scale. When SNR is high enough (greater than 3), the same scale is selected by all the

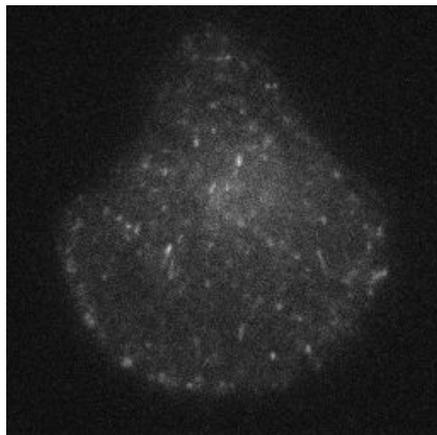
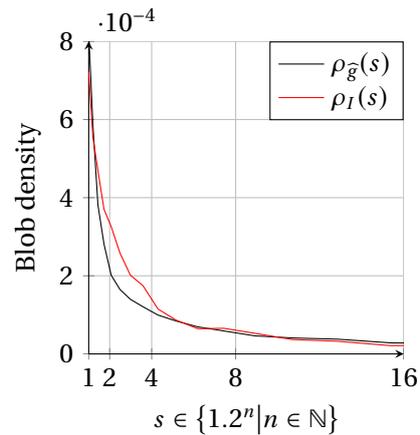
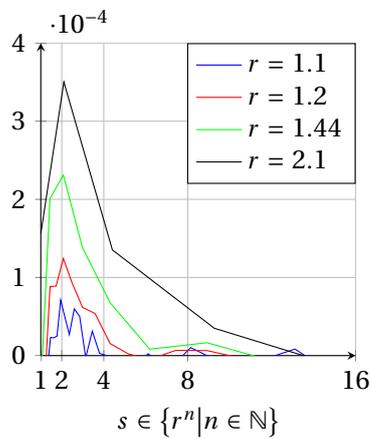
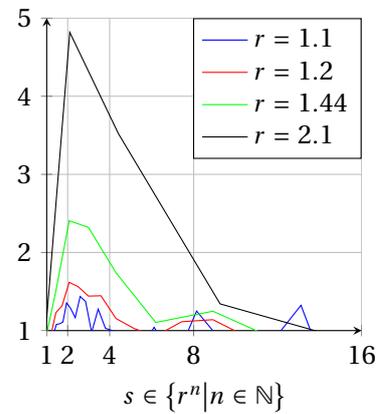
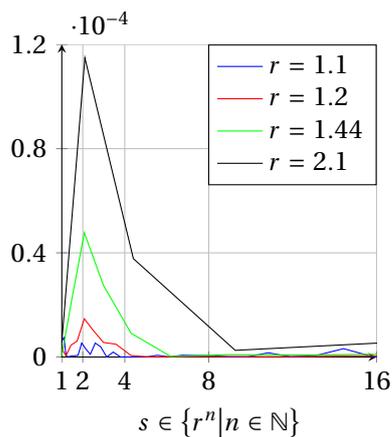
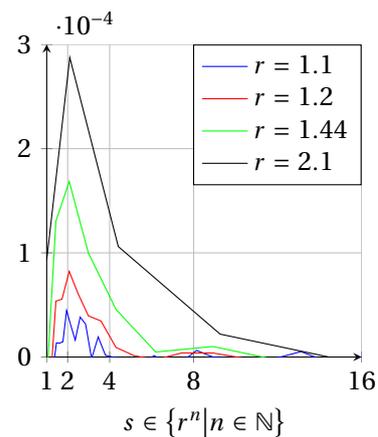
(a) Input TIRFM image I (b) Blob density in the Gaussian noise image \hat{g} and in the input image I of (a)(c) Criterion C_D (d) Criterion C_R (e) Criterion C_P (f) Criterion C_{KL}

Figure 4.2 – Statistics on the number of blobs and behavior of the scale selection criteria. (a) Input TIRFM image. (b) Distribution of the number of blobs with respect to scale for the TIRFM image and a noise image. (c,f) Output of the four scale selection criteria. In all cases, the maximum is reached for $s^* \approx 2$ in this example.

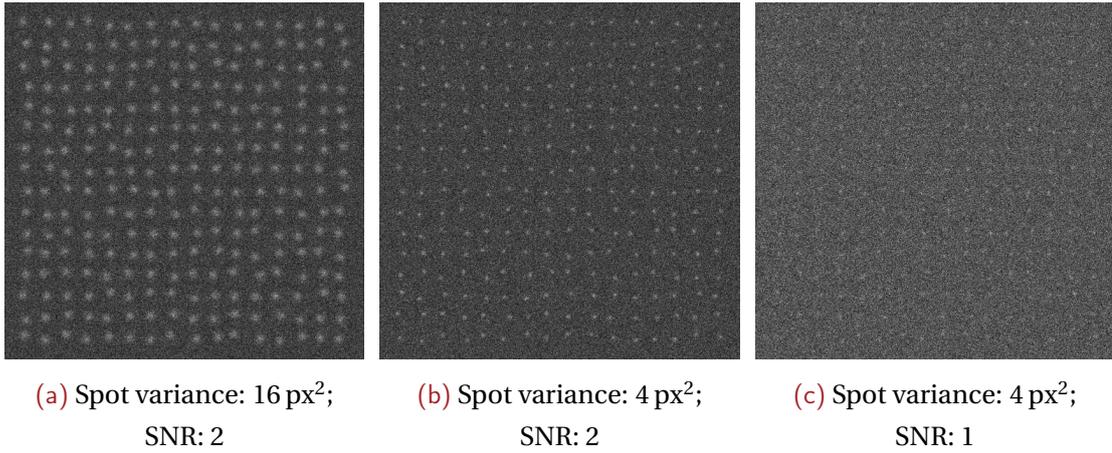


Figure 4.3 – Synthetic images with Gaussian spots corrupted by a Poisson noise

proposed criteria. The selected scale is different however, when objects are small and noise level high. In this case, C_D , C_P and C_{KL} tend to underestimate the characteristic scale, while C_R tends to overestimate it. If a scale lower than the spot variance is selected, the image is less filtered and less irrelevant structures are smoothed out, then, more false detections arise. On the contrary, a scale slightly higher than the spot variance is selected by C_R , so that a larger kernel further reduces noise, resulting in fewer false detections. As a consequence, for detection purpose, we will prefer C_R for higher precision. However, if a post-processing step is affordable to eliminate erroneous detections as in [Basset et al., 2014b], C_D and C_{KL} may be preferable for higher recall, since less spots will be smoothed out.

Table 4.1 – Scale selected by five criteria on synthetic images

SNR	Spot variance	Selection criterion				
		SLT-LoG	C_D	C_R	C_P	C_{KL}
1	4.00	3.00	2.99	6.19	1.00	2.99
2	4.00	3.00	3.58	5.16	1.00	3.58
3	4.00	3.00	4.30	4.30	4.30	4.30
4	4.00	3.00	4.30	4.30	4.30	4.30
5	4.00	3.00	4.30	4.30	4.30	4.30
5	1.00	3.00	1.44	1.44	1.73	1.44
5	2.25	3.00	2.49	2.99	2.49	2.49
5	4.00	3.00	4.30	4.30	4.30	4.30
5	9.00	9.00	8.92	8.92	8.92	8.92
5	16.00	9.00	15.41	15.41	15.41	15.41
5	25.00	27.00	22.19	22.19	22.19	22.19
5	36.00	27.00	31.95	31.95	31.95	31.95

In the approach we proposed in [Basset et al., 2014a], the scale ratio between successive scales was constrained to be an odd integer (3 in practice). In [Basset et al., 2014a], we also set $s_0 = 1$ as the finest scale. Choosing another value for s_0 could have helped obtaining a scale closer to the true spot variance for *some* variance, but the overall accuracy of the scale selection would not be increased. Really getting an arbitrary accuracy, as with the proposed method ATLAS, would have required to first automatically select the optimal s_0 value, which depends on the processed image, thus being an open issue by itself. For example it would be impossible to get an accurate scale both for an image containing spots of variance 9 and for an image containing spots of variance 16, with the same s_0 . In contrast, as shown in Table 4.1, with our new scale selection method and still with $s_0 = 1$, the first image would be correctly processed at scale 8.92, and the second image at scale 15.41.

5

Spot segmentation

ONCE THE OBJECT SCALE IS DETERMINED, we can proceed to actual spot segmentation. Since the scale selection step relies on LoG, it is natural to detect vesicles based on this particular filter. Furthermore, it has been shown in [Sage et al., 2005] that LoG is close to the optimal filter in applications like ours, that is detecting subresolved objects in fluorescence microscopy images.

5.1 Adaptive thresholding

As explained in Section 4.2, our goal is to extract the lowest values of the selected LoG map $H(\cdot, s^*)$. When the background is complex or the image exhibits large contrast variations, the use of a global threshold τ_H is not satisfactory, as illustrated in Figure 5.1. Instead, we propose to locally infer a threshold $\tau_H(p)$ for every point $p \in \Omega_I$

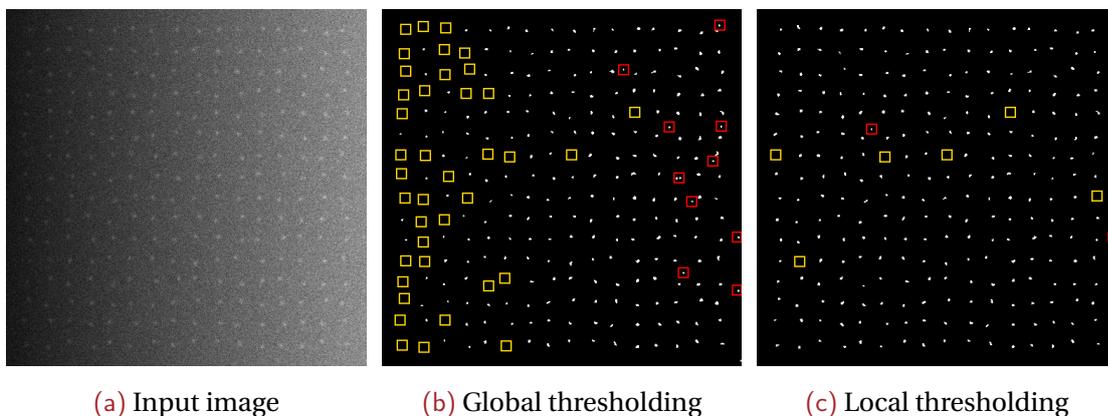


Figure 5.1 – Segmentation maps obtained with global and local thresholding. (a) Gaussian spots are added to a varying background so that contrast increases from left to right. (b) With a global threshold, segmentation maps contain both false positives (red) and false negatives (yellow). (c) With a locally adapted threshold, far better performance is achieved.

from local image statistics. To this end, we assume that the distribution of the image background is smooth and corrupted by white Gaussian noise. It holds because low frequency background structures are locally constant if the neighborhood is small enough, while noise is supposed to be normally distributed. Then, H is obtained by finite convolution of I , so that this assumption also holds for H . For every point $p \in \Omega_I$, the local mean μ_H and variance σ_H^2 are estimated over a window W_p centered in p . Then, we can infer the likelihood $\psi(p)$ of the background model $\mathcal{N}(\mu_H, \sigma_H)$ given $H(p, s^*)$:

$$\psi(p) = \Phi\left(\frac{H(p, s^*) - \mu_H}{\sigma_H}\right), \quad (5.1)$$

where Φ is the Gaussian probability cumulative density function.

Equation (5.1) can be inverted to get a threshold value below which a point is detected, according to a user-selected probability of false alarm P_{FA} , or p -value:

$$\begin{aligned} \tau_H : \Omega_I &\rightarrow \mathbb{R} \\ p &\mapsto \sigma_H \Phi^{-1}(P_{FA}) + \mu_H . \end{aligned} \quad (5.2)$$

Let us point out that we need to compute Φ^{-1} only once.

The local thresholding can thus automatically adapt to the local image statistics, while the PFA setting does not depend on the image intensity range. As a consequence, the spot detection is not affected by photobleaching when processing fluorescence microscopy image sequences. Indeed, the PFA is a parameter which is not directly related to the image properties but to the desired performance of the algorithm. Thus, it can be set once for all for a whole set of images in a given experiment. In contrast, most aforementioned detection methods have at least one threshold parameter, which directly depends on the image characteristics and has to be manually set.

5.2 Estimation window

Generally, a square window of given radius r_W centered at point $p \in \Omega_I$ is used [Kapur et al., 1985, Sahoo et al., 1988, Sahoo et al., 1997, Phansalkar et al., 2011, Sezgin and Sankur, 2004, Wilkinson and Schut, 1998]. However, while simple and fast, this type of window presents several drawbacks. First, the choice of the window size is critical due to the discontinuity at the window border; increasing or reducing the window radius by one pixel only can affect the estimated mean and variance. For the same reason, the resulting threshold map τ can be spoiled, exhibiting strong blocky artifacts as illustrated in Figure 5.2b. Also, the square window is not invariant to image rotation.

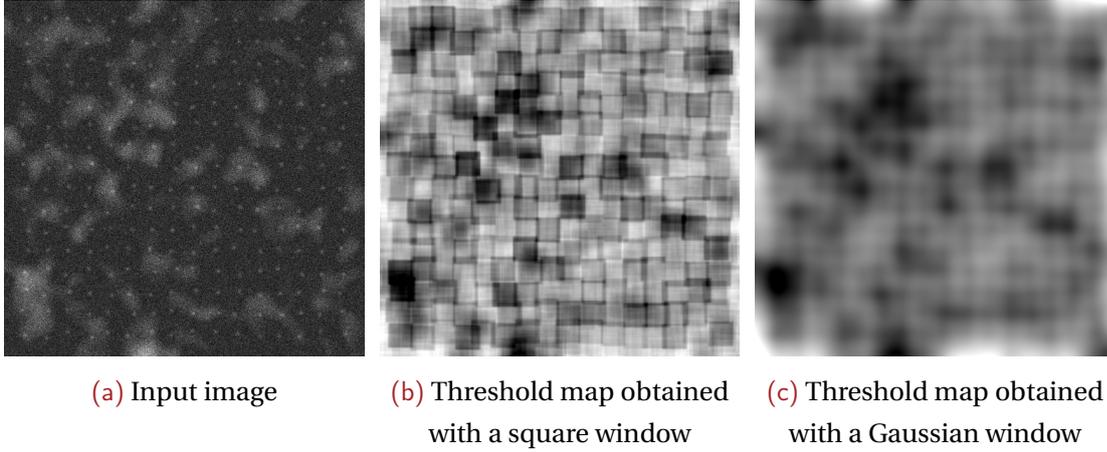


Figure 5.2 – Threshold maps with different windows. (a) Input image. (b) Use of a square window. (c) Use of a Gaussian window. Bright areas of the input image should correspond to low thresholds as the LoG is a negative filter. Strong blocky artifacts appear with the square window and background statistics are better estimated with the Gaussian window.

To cope with these undesirable effects, we prefer to utilize a Gaussian weighted window W_p with a bandwidth σ_w :

$$\forall q \in \Omega_I, W_p(q) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{\|q - p\|_2^2}{2\sigma_w^2}\right). \quad (5.3)$$

To speed up the statistics estimation, we rely on the recursive Deriche filter [Deriche, 1990]. The good properties of the Gaussian window are kept even if the window profile is not strictly Gaussian. As reported in the next section, the smooth weight decrease allows for a very low sensitivity of the window size parameter. Furthermore, giving more importance to central points naturally yields a better estimation of local statistics. Figure 5.2 demonstrates the advantage of such a smooth windowing compared to the square one. The bright background structures of the input image are well reflected in the threshold map of Figure 5.2c, while large errors are encountered in Figure 5.2b for the square window.

The computation load with Gaussian windows will obviously be higher than with square patches, but not dramatically. Tests were performed on a laptop with 2.3 GHz Intel i7 processor. For 512×512 images, the computation time is 50 ms per image with a quasi-Gaussian window, and 15 ms with a square one relying on a moving average algorithm [Kenney and Keeping, 1947, Crow, 1984, Shafait et al., 2008]. For both windows, the numerical complexity is linear with the size of the image.

Finally, in an image sequence, frames are segmented independently, so that the program could even be sped up by processing the frames in parallel. Let us also stress that, if the image background is stationary, the estimation window could be extended

to a spatiotemporal window in order to estimate the background statistics still more robustly.

5.3 Experimental results

We have compared ATLAS to state-of-the-art spot detection methods in a wide variety of cases. Comparative quantitative evaluation was carried out on several datasets with ground-truth. The first dataset is generated with the *Synthetic Data Generator* [Image] plugin introduced in [Smal et al., 2010] and later used in [Basset et al., 2014a, Rezatofghi et al., 2012]. Twelve methods were compared on this dataset, which is (to our knowledge) the most complete comparison of spot detection methods to date, but the images remain somewhat too artificial and too simple.

As mentioned in the Introduction section, we have conducted comparative experiments on three other datasets involving more complex contents with the most competitive detection methods, namely MS-VST, MPHD, HD and conditional random fields for protein transport carriers segmentation (C-CRAFT). First, Boulanger et al. [Boulanger et al., 2009] and Rezatofghi et al. [Rezatofghi et al., 2013] proposed particle dynamics simulators, referred in the sequel as *Traffic simulator* and *TIRFM simulator*, respectively. The *Traffic simulator* was used in [Pécot et al., 2015] to evaluate the performance of several methods. Secondly, we have constructed another image dataset named *Spot in M10* where image backgrounds are extracted from real TIRFM images.

As stated in Section 2.3.4, all the images processed in this section are first stabilized using the generalized Anscombe transform (GAT)-based variance stabilization method described in [Boulanger et al., 2010b].

5.3.1 Performance measures

ATLAS delivers a binary detection map. In order to evaluate the performance of the method and compare it to other ones, we compute the centroid of every segmented connected component, resulting in a set of locations $\{\delta\}$. Then, following [Smal et al., 2010], an object ω of the ground-truth is correctly detected if and only if: (1) its nearest neighbor δ in the set of detected centroids is closer than 4 pixels away, and (2) ω is also the nearest neighbor of δ in the ground-truth set of locations. Let us denote N_{TP} the number of true positives, N_{FP} the number of false positives and N_{FN} the number of false negatives.

We can evaluate different scores for every image and parameter setting. As in [Basset et al., 2014a, Rezatofghi et al., 2012, Smal et al., 2010], we compute the true positive ratio $TPR = N_{TP}/(N_{TP} + N_{FN})$ and the modified false positive ratio $FPR^* = N_{FP}/(N_{TP} + N_{FN})$. The value of TPR when $FPR^* = 0.01$ is denoted TPR^* and is used

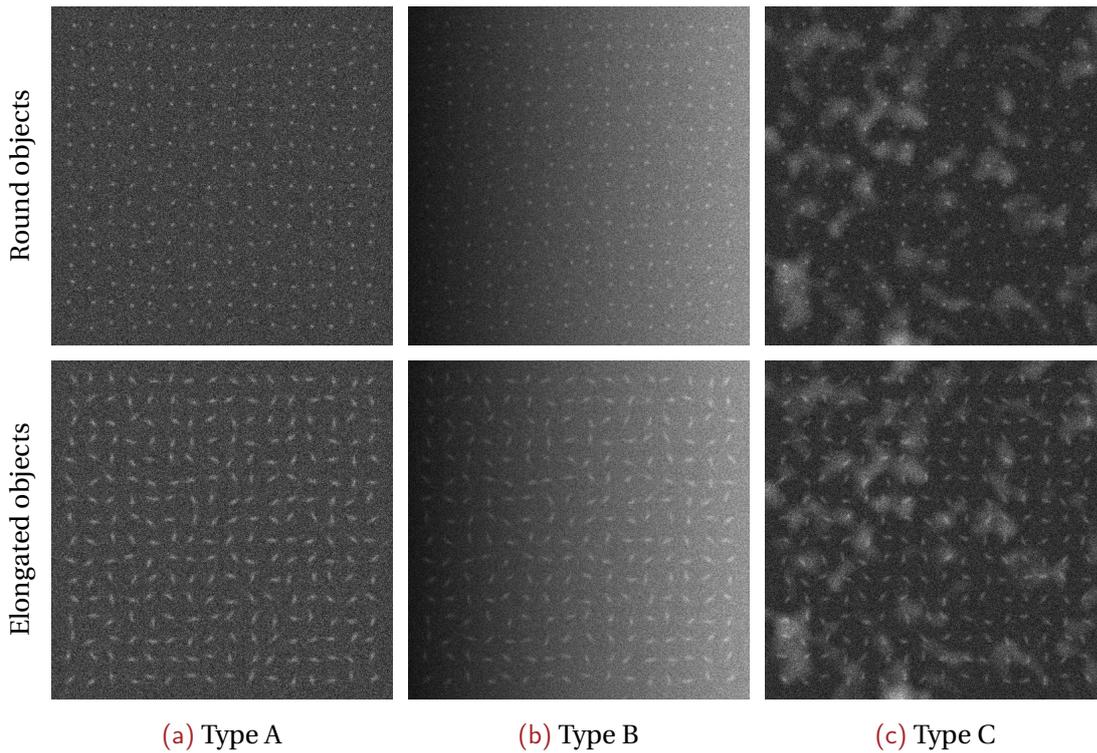


Figure 5.3 – Sample images from the Synthetic Data Generator benchmark for SNR = 2. Types are defined in the main text.

to compare methods in [Basset et al., 2014a, Rezatofighi et al., 2012, Smal et al., 2010]. Moreover, to compare ATLAS with the detection methods tested in [Pécot et al., 2015], namely HD, MS-VST and C-CRAFT, we compute the precision $\text{Prec} = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FP}})$ and recall $\text{Rec} = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FN}})$. Varying the threshold parameter for the existing methods or the PFA value for ATLAS, we can plot the free receiver-operator characteristic (FROC), that is the TPR-versus-FPR* curve, and the precision-versus-recall curve. That way, the behaviors of the methods can be evaluated more thoroughly. Additionally, we compute the area under the FROC curve as a performance score over a wide range of thresholds or PFA values. We also resort to the F-measure defined by the harmonic mean of precision and recall $F = 2 \text{Prec} \cdot \text{Rec} / (\text{Prec} + \text{Rec})$, and more precisely to the best reachable F-measure F^* .

5.3.2 Synthetic Data Generator

In [Basset et al., 2014a, Rezatofighi et al., 2012, Smal et al., 2010], twelve methods are evaluated over six image sets of 16 images each. They are depicted in Figure 5.3. Two object shapes are considered: isotropic Gaussian spots of standard deviation 2 pixels, and elliptic Gaussian spots of standard deviations 5 and 2 pixels along the two principal axes, respectively. Three types of background are generated: uniform intensity (type A), horizontal intensity gradient (type B), and large random structures (type C).

Table 5.1 – Comparison of TPR* values with state-of-the-art methods on the Synthetic Data Generator dataset for SNR = 2

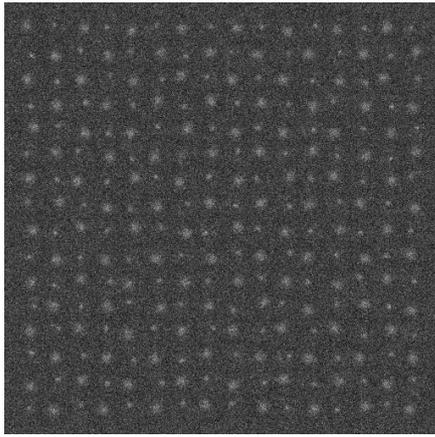
Object shape Background type		Round			Elongated		
		A	B	C	A	B	C
TPR*s of supervised detection methods							
AB	[Jiang et al., 2007]	0.99	0.94	0.94	0.99	0.99	0.99
FDA	[McLachlan, 1992]	0.99	0.99	0.96	0.99	0.99	0.99
TPR*s of unsupervised detection methods							
TH	[Bright and Steel, 1987]	0.99	0.88	0.48	0.99	0.96	0.56
HD	[Smal et al., 2008]	0.99	0.97	0.90	0.99	0.99	0.97
WMP	[Olivo-Marin, 2002]	0.81	0.37	0.30	0.31	0.17	0.18
IFD1	[ter Haar Romeny, 2003]	0.98	0.67	0.89	0.53	0.31	0.31
IFD2	[ter Haar Romeny, 2003]	0.99	0.46	0.71	0.59	0.23	0.19
MTH	[Soille, 2003]	0.99	0.87	0.88	0.99	0.98	0.91
SEF	[Sage et al., 2005]	0.99	0.91	0.95	0.99	0.99	0.95
MS-VST	[Zhang et al., 2007]	0.99	0.99	0.93	0.99	0.99	0.96
MPHD	[Rezatofighi et al., 2012]	0.99	0.97	0.94	0.99	0.99	0.95
SLT-LoG	[Basset et al., 2014a]	1.00	0.99	0.98	1.00	1.00	1.00
ATLAS		1.00	1.00	0.99	1[†]	1[†]	1.00

[†]1 means that every spot is correctly detected, while 1.00 corresponds to a rounded value

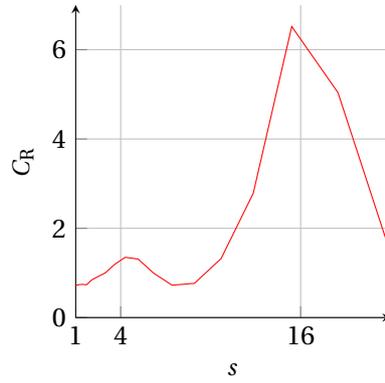
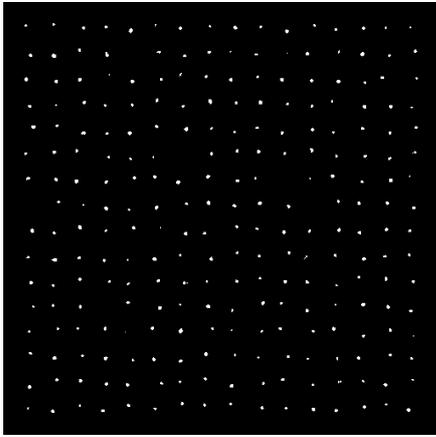
A Poisson noise is added to obtain a SNR of 2. Table 5.1 summarizes the TPR* values obtained by the detection methods on this benchmark. Number 1.00 corresponds to rounded values, while 1 means that every spot is correctly detected. In the later case, the 4096 spots of the sequence are all recovered even for a FPR* value lower than 0.01.

In [Rezatofighi et al., 2012, Smal et al., 2010], in order to perform a fair comparison, all the methods were run with various parameter settings and, for each method, the setting yielding the best performance was selected, that is the highest TPR for FPR* = 0.01. Therefore, reported results reflect the best possible performance of the compared methods. As for ATLAS, we have fixed the size of the background estimation window σ_W once for all to 15 pixels, that is all the sequences were segmented using the same window parameter. Notwithstanding, as demonstrated in Table 5.1, ATLAS performs better than any other method described in [Basset et al., 2014a, Rezatofighi et al., 2012, Smal et al., 2010] on every image subset of the benchmark dataset.

Finally, we have conducted a complementary experiment to demonstrate the ability of the method to detect two different scales in an image. While not relevant in our study, the ability to deal with images containing objects of different sizes could be appealing in some other applications. Therefore, a series of 16 synthetic images containing 2048 spots of variance 4 px² and 2048 spots of variance 16 px² was generated. A Poisson noise was added to get a SNR of 2 (which converts to 6 dB). The C_R criterion



(a) Input image

(b) Output of criterion C_R 

(c) Two-scale segmentation results: map and number of spots detected

	Spot variance		FPR*
	4 px ²	16 px ²	
	TPR		
$\hat{s}_1 = 4.29$	0.98	0.58	0.01
$\hat{s}_2 = 15.41$	0.14	0.97	0.00
Maps union [†]	0.98	0.98	0.01

[†]The maps union is obtained by combining the two segmentation maps (a point is detected in the union if it is detected in the first or in the second map).

Figure 5.4 – Case of images containing spots of two different sizes. (a) Input image with spots of two different sizes (resp. 4 and 16 px²). (b) Two local maxima are located by C_R at about 4 and 16 px², corresponding to the spots variances. (c) Segmentation map when the LoG filter scale is set to the first local maximum in C_R , that is $\hat{s}_1 = 4.29$, then to the second local maximum $\hat{s}_2 = 15.41$.

output is plotted in Figure 5.4. It is computed over the 16 images. It exhibits two local maxima at scales $\hat{s}_1 = 4.29$ and $\hat{s}_2 = 15.41$, so that two characteristic scales could be detected as well if specified by the user. We have computed the segmentation maps for the two maximal scales. As shown in Figure 5.4, most spots of variance 4 are detected at scale 4.29, while most spots of variance 16 are detected at scale 15.41. Also, the union of both segmentation maps yields convincing detection results for such a multiscale case, that is TPR = 0.98 for FPR* = 0.01. The detection of spots of different sizes is not an objective of the thesis, but this experiment shows that our method can be amenable to detect spots of several sizes. Besides, this experiment demonstrates the necessity of finding the right LoG scale. Indeed, the scale 4.29 is not adequate to detect the spots of variance 16 (with a TPR* of 0.58) and the same holds for detecting spots of variance 4 at scale 15.41 (TPR* = 0.14).

Table 5.2 – Comparison with LR-MRF on the Subcell dataset

	Prec	Rec	F-score
LR-MRF	0.876	0.858	0.867
ATLAS	0.994	0.957	0.969

As for the supervised method logistic regression with Markov random field (LR-MRF) described in [Ruusuvuori et al., 2012], the authors reported better results than those of 11 unsupervised methods on a synthetic dataset of 20 images named *Subcell* [Ruusuvuori et al., 2010]. In addition, we have evaluated our method on the same synthetic dataset. We give the results in Table 5.2, which demonstrates that our method ATLAS also outperforms the 12 supervised method of [Ruusuvuori et al., 2012].

5.3.3 Spot in M10 dataset

To further challenge spot detection methods, we have generated another synthetic image dataset to carry out comparative objective evaluation. The so-called *Spot in M10* dataset was constructed by mixing real background images and synthetic spots, as illustrated in Figure 5.5. To obtain realistic backgrounds, the rolling-ball extraction method [Sternberg, 1983] was first applied to one hundred real TIRFM images. Other background extraction methods could have been used as well. Actually, we do not need a very precise estimation of the background to construct the dataset. Furthermore, in order to ensure a fair comparative evaluation, it is preferable to use a background extraction method which is not part of any spot detection methods. Then, for each background image, 45 to 80 vesicles, modeled as Gaussian spots, were randomly added inside the cell region. Two spot variance values were tested: 1 px^2 and 1.44 px^2 . Finally, the resulting images are corrupted by various Poisson-Gaussian noises to form images very similar to real ones.

We prefer to rely on the peak signal-to-noise ratio (PSNR) instead of SNR to evaluate the noise level in fluorescence microscopy images. Indeed, it is more straightforward and easier to compute PSNR in real images. After stabilizing the variance, the PSNR can be estimated as the difference between the minimum and maximum intensities, since the noise variance after stabilization is supposed to equal 1. The estimated PSNR of the real TIRFM images we are dealing with ranges from 23 to 31 dB. In order to evaluate the robustness to noise of the methods, we have generated even noisier images, down to PSNR = 18 dB. Note that PSNRs are higher than aforementioned SNRs since we are considering the brightest spot in the sequence. As shown in Figure 5.5, we end up with a set of six image collections, each containing

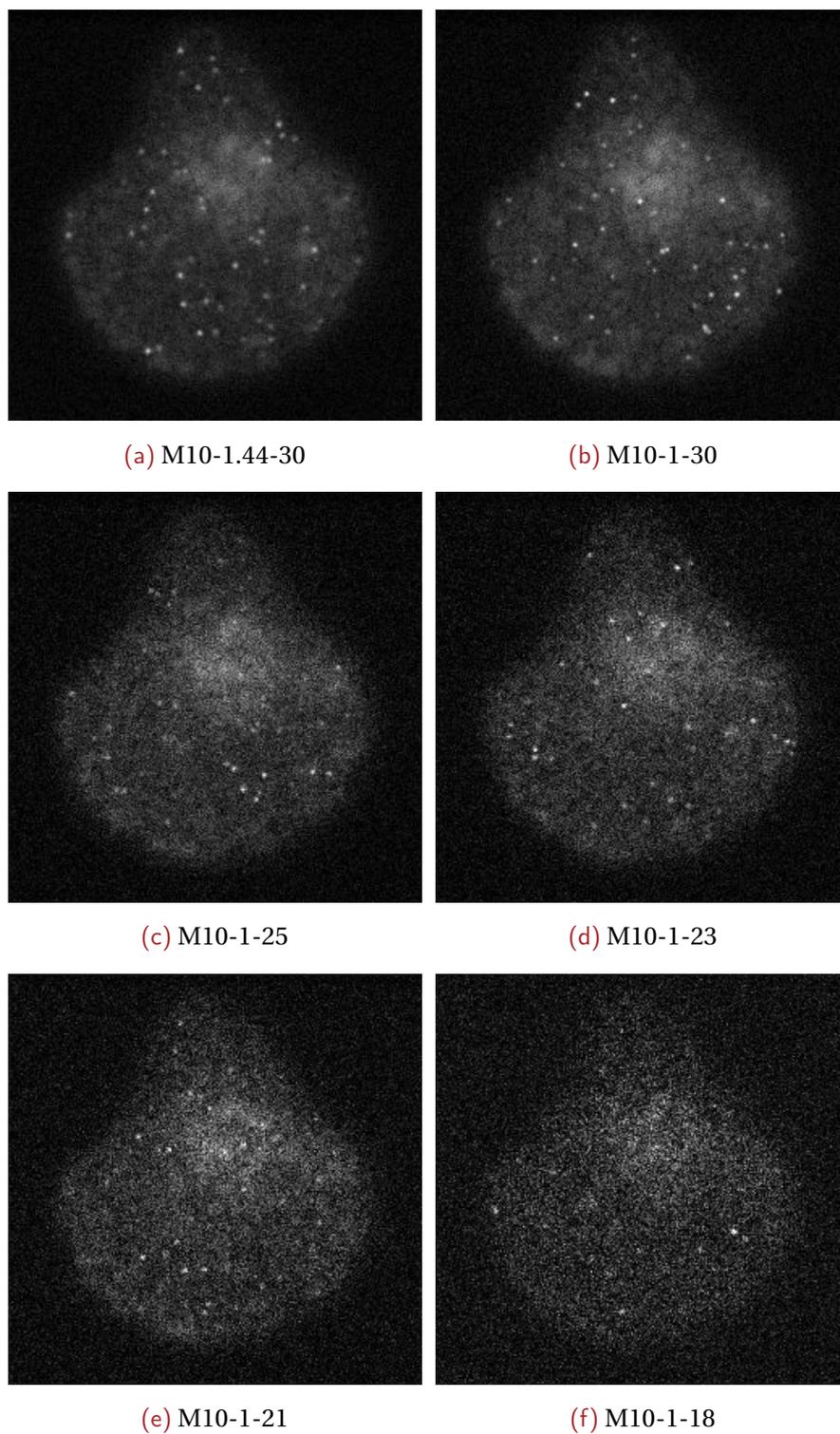


Figure 5.5 – (a) Constructed image containing spots of variance 1.44 px^2 ; estimated PSNR: 30 dB. (b-f) Constructed images containing spots of variance 1 px^2 ; PSNR decreases from 30 dB (b) to 18 dB (f).

Table 5.3 – Comparison with MS-VST and MPHD on the Spot in M10 dataset

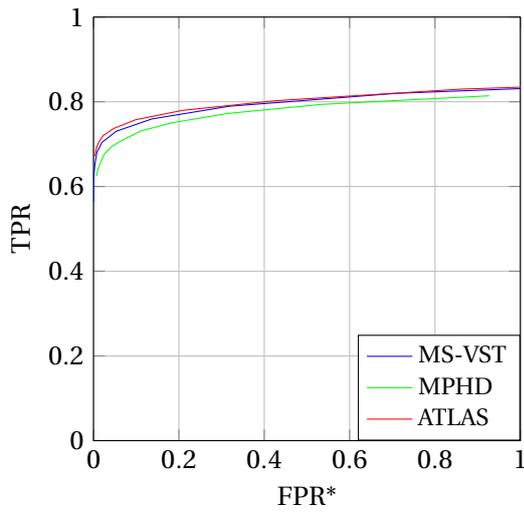
Spot variance	1.44	1	1	1	1	1
PSNR	30	30	25	23	21	18
TPR at FPR* = 0.01						
MS-VST	0.69	0.71	0.46	0.40	0.29	0.13
MPHD	0.64	0.66	0.50	0.44	0.35	0.18
ATLAS	0.70	0.71	0.50	0.45	0.34	0.18
Area under FROC curve						
MS-VST	0.79	0.80	0.65	0.60	0.52	0.36
MPHD	0.77	0.78	0.66	0.61	0.53	0.39
ATLAS	0.80	0.80	0.69	0.64	0.55	0.44
Maximum F-score F*						
MS-VST	0.82	0.83	0.68	0.64	0.56	0.41
MPHD	0.80	0.81	0.70	0.66	0.58	0.44
ATLAS	0.83	0.83	0.71	0.67	0.58	0.48

one hundred images and more than six thousand spots. The dataset is available at <https://serpico.rennes.inria.fr>.

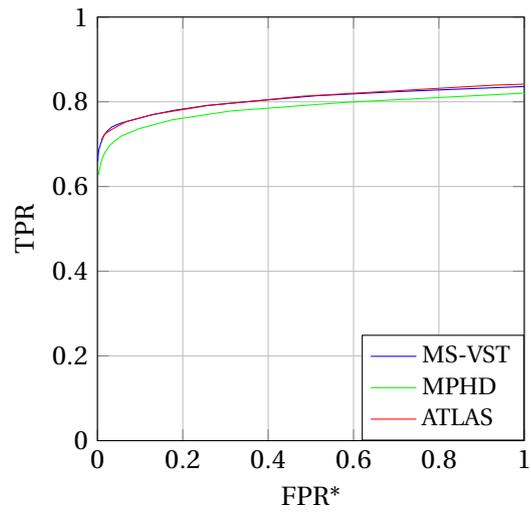
FROC curves are plotted in Figure 5.6 for each image collection. MS-VST performed better than MPHD at high PSNR, while the opposite holds for noisier images. Our method ATLAS is on par or slightly better than MS-VST at high PSNR, and demonstrates better robustness to noise. Therefore, FROC curves obtained with ATLAS nearly always dominate others. Table 5.3 summarizes TPR* scores, areas under FROC curves and best F-scores F* of the three methods. Our method always ranks first, except in one case, where MPHD is slightly better around FPR* = 0.01. Whatever the evaluation measure, the best overall performance is achieved by ATLAS.

The sensitivity of the parameters involved in MS-VST and ATLAS was low enough to keep the same setting for the whole dataset. For MS-VST, scale levels 2 and 3 were used. For ATLAS, the window radius σ_W was set to 60 pixels. Moreover, as illustrated in Figure 5.7, this window size is not critical. To achieve the best performance with MPHD, we varied the LoG standard deviation from 0.83 pixel for highest PSNR to 1 pixel for lowest PSNR. Results were unchanged when varying the maximum dome radius (Table 3.1) from 5 to 20 pixels.

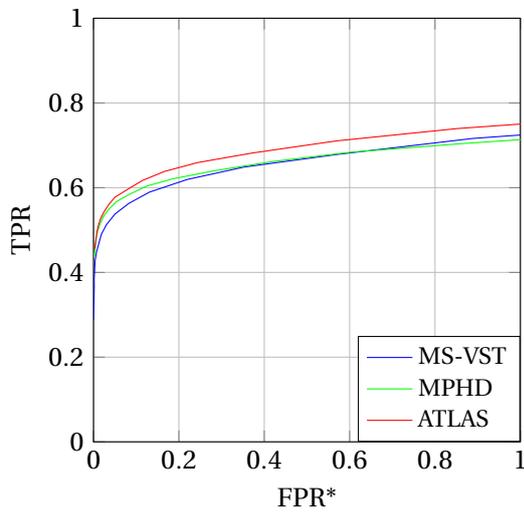
To further investigate the behavior of ATLAS when varying the parameters, we have conducted additional experiments. First, Figure 5.7 demonstrates the extremely low sensitivity of the estimation window size. FROC curves are indistinguishable when doubling σ_W from 40 to 80 pixels. Secondly, we evaluated the precision of the scale selection as reported in Figure 5.8. In the M10-1-30 image subset, spots have a variance of 1 px^2 , so that one might think that best results would be achieved with a matched LoG filter of variance 1 px^2 . Yet, the scale selected with criterion C_R is higher,



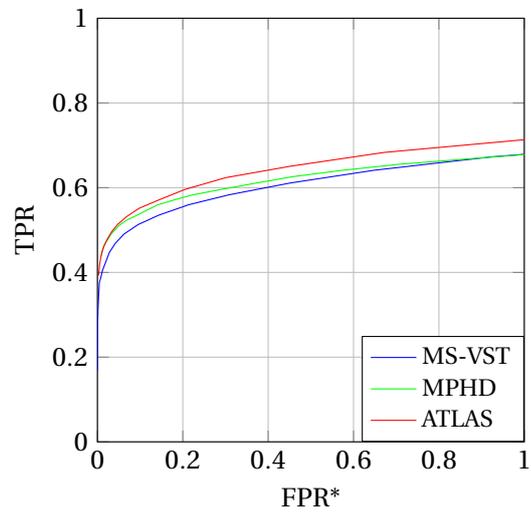
(a) Spot variance: 1.44; PSNR: 30 dB



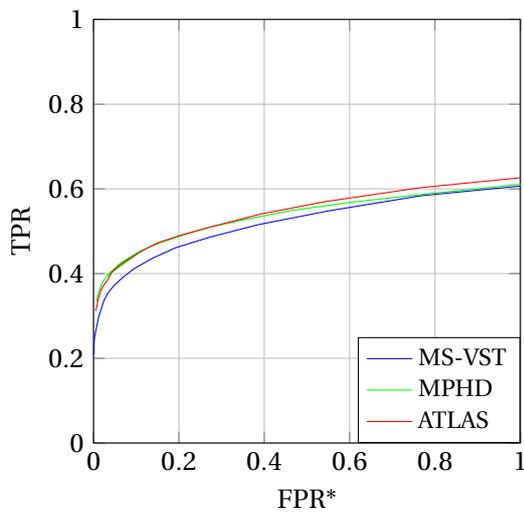
(b) Spot variance: 1; PSNR: 30 dB



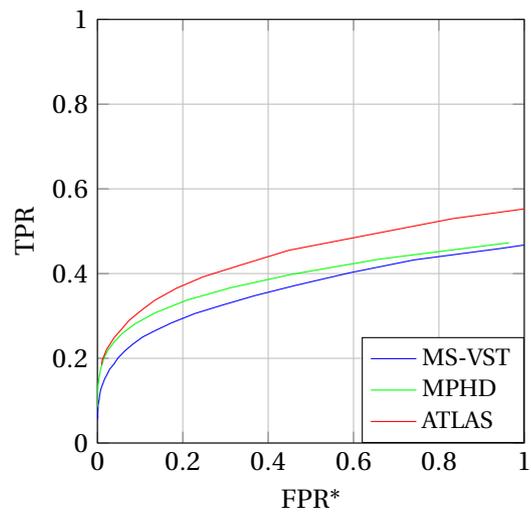
(c) Spot variance: 1; PSNR: 25 dB



(d) Spot variance: 1; PSNR: 23 dB



(e) Spot variance: 1; PSNR: 21 dB



(f) Spot variance: 1; PSNR: 18 dB

Figure 5.6 – Comparison of FROC curves obtained with ATLAS, MPHD and MS-VST on the proposed *Spot in M10* dataset

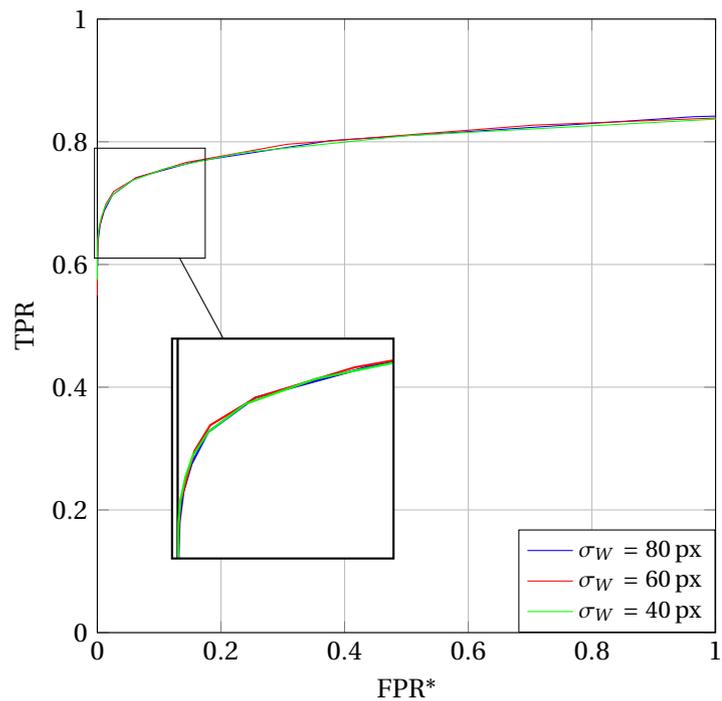


Figure 5.7 – Sensitivity of the window parameter of ATLAS on the M10-1-30 image collection. Doubling the Gaussian window size does not impact performance.

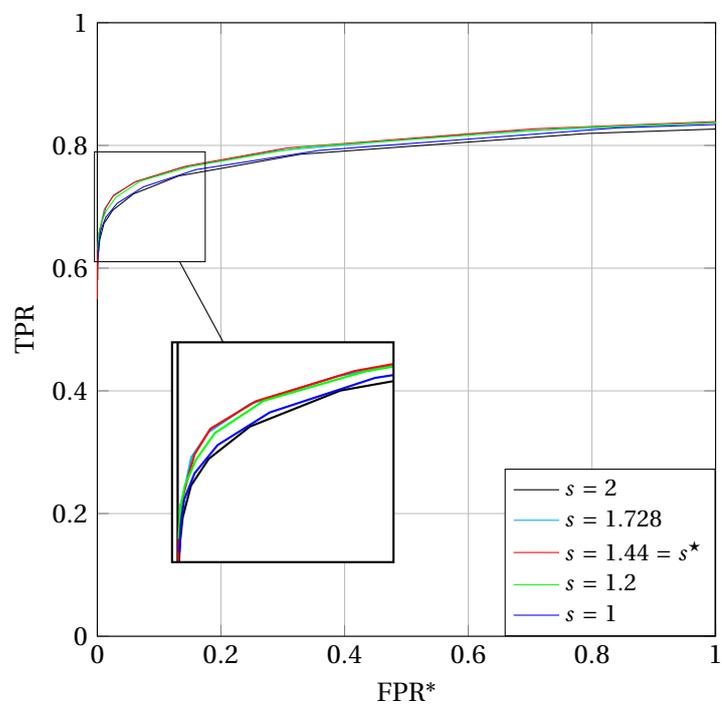


Figure 5.8 – Precision of the scale selection illustrated on the M10-1-30 image collection. Performance was evaluated by taking different scales in the LoG filter. Best results are obtained with the automatically selected scale $s^* = 1.44$.

namely 1.44 px^2 . In order to check the relevance of the selected scale, we run ATLAS with several LoG variances s taken in $\{1, 1.2, 1.44, 1.73, 2\}$. For the sake of visibility, we only plotted results with $s \in \{1, 1.44, 2\}$ in Figure 5.8. The FROC curve obtained with the scale 1.44 px^2 corresponding to s^* dominates others, which demonstrates the actual relevance of the selected scale and the efficiency of ATLAS. Besides, this shows the benefit of applying the LoG at the appropriate scale, and justifies the proposed method.

5.3.4 TIRFM simulator

In [Rezatofighi et al., 2013], the authors propose a method for simulating realistic TIRFM images containing spots undergoing a Brownian motion. The background of a real TIRFM image is first extracted using the h -dome method MPHD [Rezatofighi et al., 2012]. Then, objects are randomly spread inside the cell and a Poisson-Gaussian noise is added. We have evaluated the performance of our method on two sequences provided by the authors (Figure 5.9a-d), and compared it to the two best ranked methods in [Rezatofighi et al., 2012, Smal et al., 2010], namely MS-VST and MPHD. FROC curves are plotted in Figure 5.9e,f for both sequences. While negligible in the first sequence, photobleaching is strong in the second one, so that all the methods performed worse in the latter case. ATLAS obtained better results than MS-VST on both sequences, and than MPHD on the second sequence. However, MPHD performed significantly better than ATLAS and MS-VST on the first sequence for low FPR*. Let us nevertheless remind that the backgrounds were extracted using MPHD.

5.3.5 Traffic simulator

We can also compare our method to a robust patch-based method named C-CRAFT, which jointly estimates background and segments vesicles in 2D or 3D fluorescence image sequences, in the framework of conditional random fields [Pécot et al., 2015]. Vesicle segmentation and background estimation are formulated as a global energy minimization problem.

Even if spot detection is only one of the goals of C-CRAFT, the authors also compared their method to several spot detectors, including HD and MS-VST, on image sequences generated with the vesicle traffic simulation method [Boulanger et al., 2009]. Therefore, we can compare the spot detection results of ATLAS on the dataset used in [Pécot et al., 2015]. It is composed of four sequences of 120 frames. Two real backgrounds were extracted for *Circle*- and *Crossbow*-shaped micropatterned cells. Vesicle motions are then simulated, and both sequences are corrupted with two different Poisson-Gaussian noises. Since no FROC curves were given in [Pécot et al., 2015],

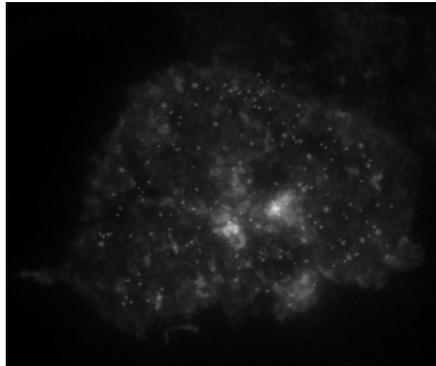
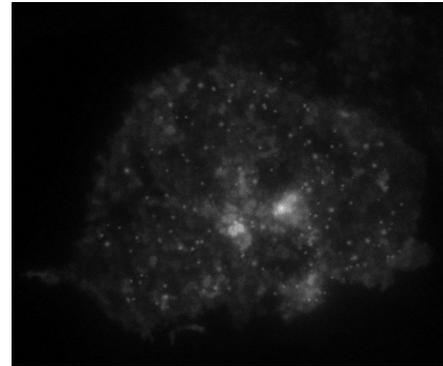
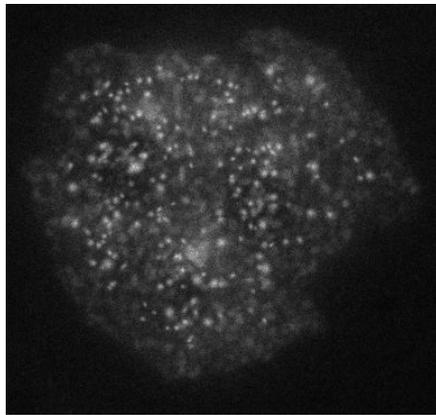
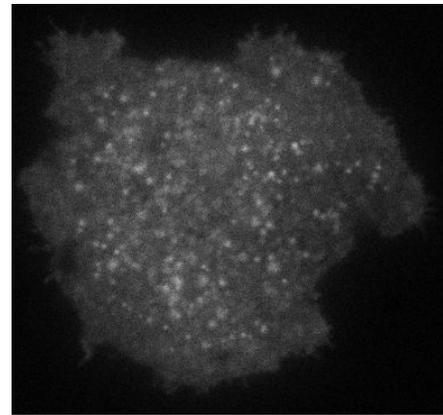
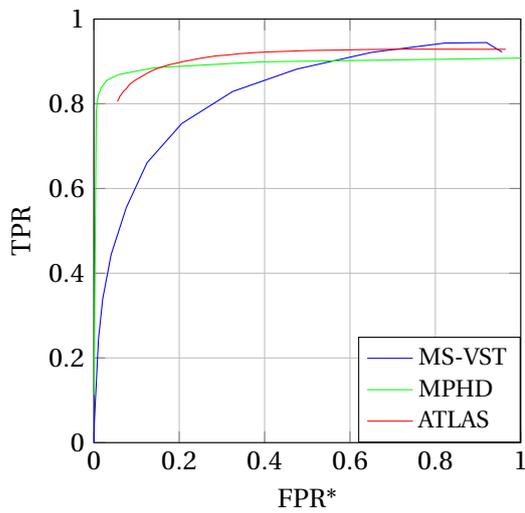
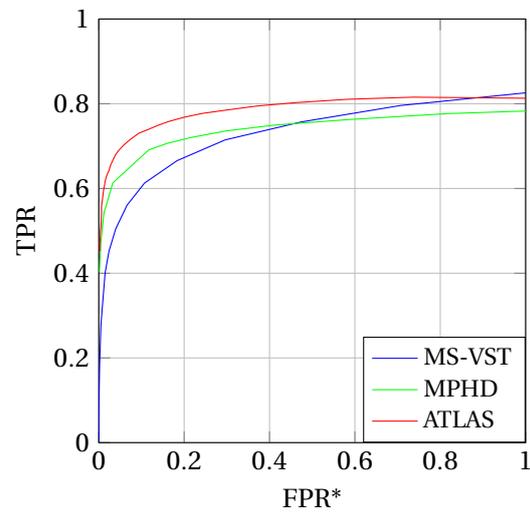
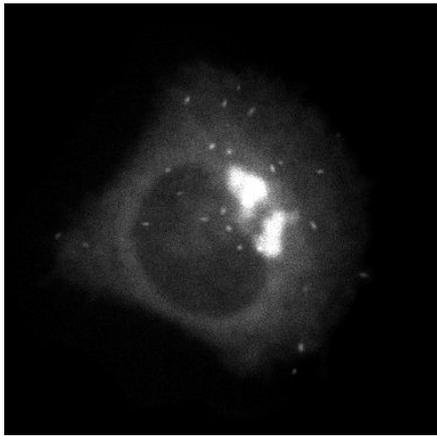
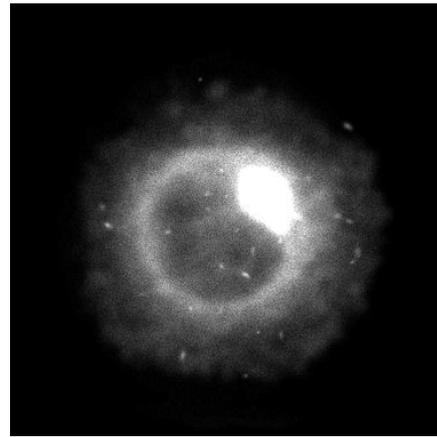
(a) First frame of the *EstBack1* sequence(b) Last frame of the *EstBack1* sequence(c) First frame of the *EstBack2DS* sequence(d) Last frame of the *EstBack2DS* sequence(e) Results on the *EstBack1* sequence(f) Results on the *EstBack2DS* sequence

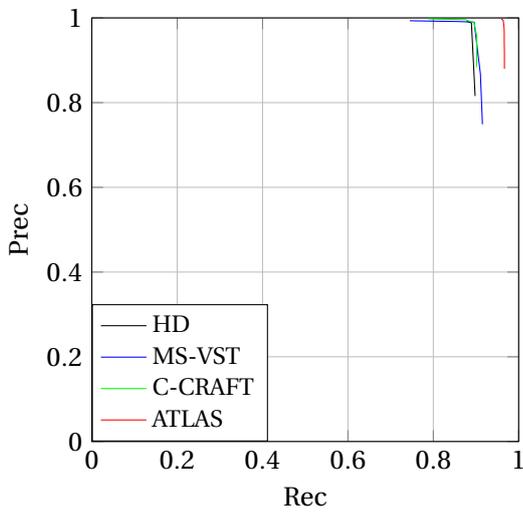
Figure 5.9 – Comparison of FROC curves obtained with ATLAS, MPHD and MS-VST on the *TIRFM simulator* dataset



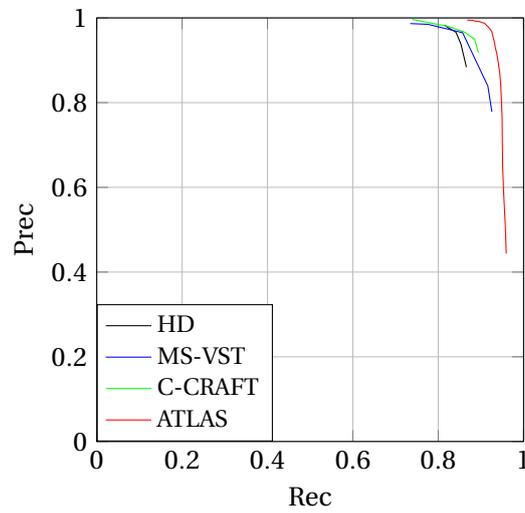
(a) First frame of the *Crossbow* sequence
(PSNR: 31 dB)



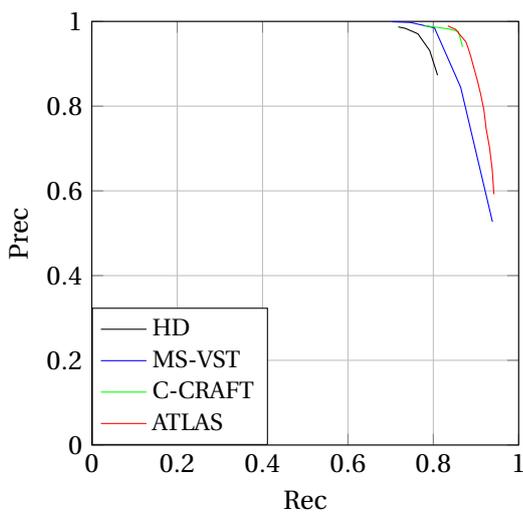
(b) First frame of the *Circle* sequence
(PSNR: 33 dB)



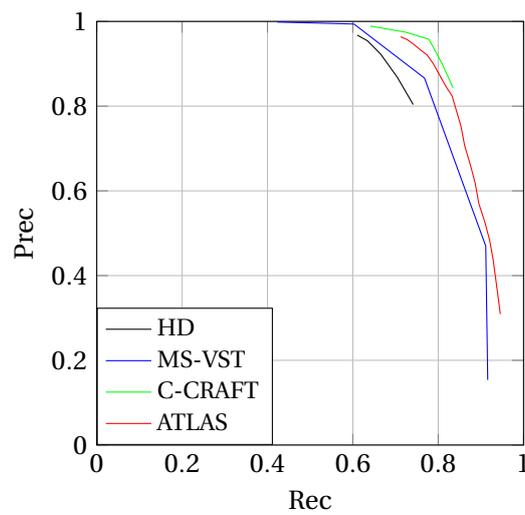
(c) *Crossbow* sequence; PSNR: 31 dB



(d) *Crossbow* sequence; PSNR: 25 dB



(e) *Circle* sequence; PSNR: 33 dB



(f) *Circle* sequence; PSNR: 31 dB

Figure 5.10 – Comparison of precision-recall curves obtained with ATLAS, C-CRAFT, HD and MS-VST on the *Traffic simulator* dataset.

precision-recall curves of ATLAS are plotted in Figure 5.10 and compared to the results provided in [Pécot et al., 2015] for C-CRAFT, HD and MS-VST.

On the two *Crossbow* sequences, ATLAS outperforms the three methods. On the *Circle* sequence, which exhibits extremely contrasted background, ATLAS and C-CRAFT are on par for the highest PSNR (33 dB), but C-CRAFT performs better when noise level is high and ATLAS ranks second. This can be explained by the temporal integration used by C-CRAFT to jointly estimate background and detect vesicles.

5.3.6 Real images

Finally, we have conducted experiments on real images. As in most experiments on microscope images, it is not possible to extract the full ground-truth, but only to get an expert annotation, which is not perfect by nature. As a consequence, a quantitative comparison would not be completely reliable. Therefore, qualitative comments are given to analyze results on real data.

We have processed two real sequences of M10 cells expressing various fluorescently tagged proteins, namely Langerin and TfR tagged with pHluorin, and Rab11 tagged with mCherry.

The first 300-frame sequence was acquired with the Dual View optical beam-splitting technique [Gidon et al., 2012]. As explained in Chapter 2, with this technique, several fluorescence channels are simultaneously acquired side-by-side on the sensor, resulting in images including two parts with very different intensity ranges, as shown in Figure 5.11a. The upper half of the image shows Rab11 GTPase. Overall, it is much brighter than the lowest half showing Langerin. Even in that complicated case, the adaptive thresholding approach produces very satisfactory results in both parts of the image. In the lower part, spots even darker than the upper part cell background are detected, while no false detection appears even in the brightest region of the upper part.

As stated in the introduction, various images fulfill the characteristic scale assumption. To show the ability of the method to cope with a very different type of images, ATLAS was also successfully applied to tissue micro-array (TMA) images, where the cores must be first detected, before being individually analyzed. In these images, the difficulty mainly comes from the high variability of the cores intensity. Indeed, as illustrated in Figure 5.12, spurious bright dots have a much higher intensity than most cores. To smooth out these dots while still detecting low-intensity cores, the LoG scale must be precisely set. Cores are correctly detected by ATLAS, and only one false positive arises corresponding to a broken but intense core. However, this kind of error can be easily discarded by post-processing [Nguyen et al., 2015].

We have also compared the binary segmentation maps supplied by ATLAS and MS-VST on a second TIRFM image sequence of 300 images. We only compared ATLAS

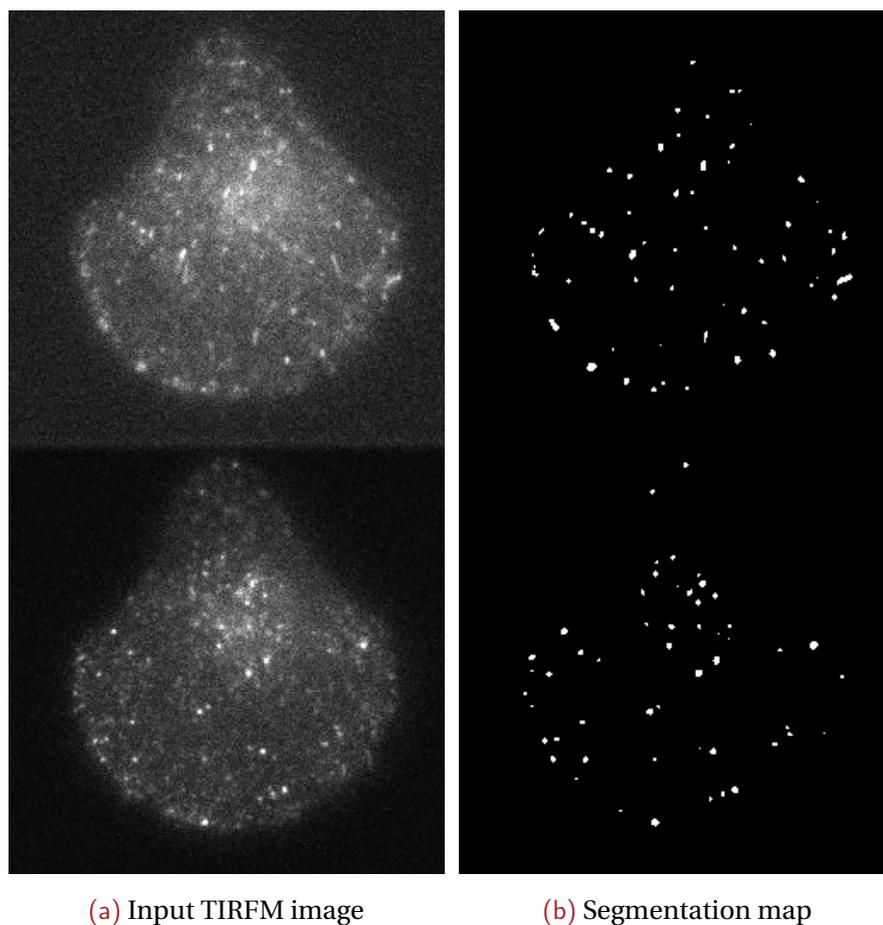


Figure 5.11 – Segmentation results supplied by ATLAS on a real two-channel TIRFM image sequence. (a) First frame of the sequence. The two channels are displayed one above the other, Rab11-mCherry channel on top, Langerin-pHluorin below. (b) Segmentation of vesicles by ATLAS.

to MS-VST because the latter was assessed to be the best competitor. As for MPHD, it only provides spot center coordinates. As shown in Figure 5.13a, elongated objects present in Rab11-mCherry sequences are better recovered by our method. This demonstrates the ability of the LoG filter to detect elongated structures, providing one dimension of the elongated structures nearly corresponds to the selected scale. The thresholding step is performed pixelwise, so that the elongated structures can be segmented by successively detecting neighboring pixels forming the elongated connected component. In this case, the accurate scale selection of ATLAS yields better segmentation results than MS-VST, in which the ratio between consecutive scales is too high (and it cannot be reduced) to correctly handle the sequence characteristic scale. It results in elongated objects of Figure 5.13a being split in the segmentation map delivered by MS-VST (Figure 5.13b), while elongated connected components are correctly recovered by ATLAS (Figure 5.13c).

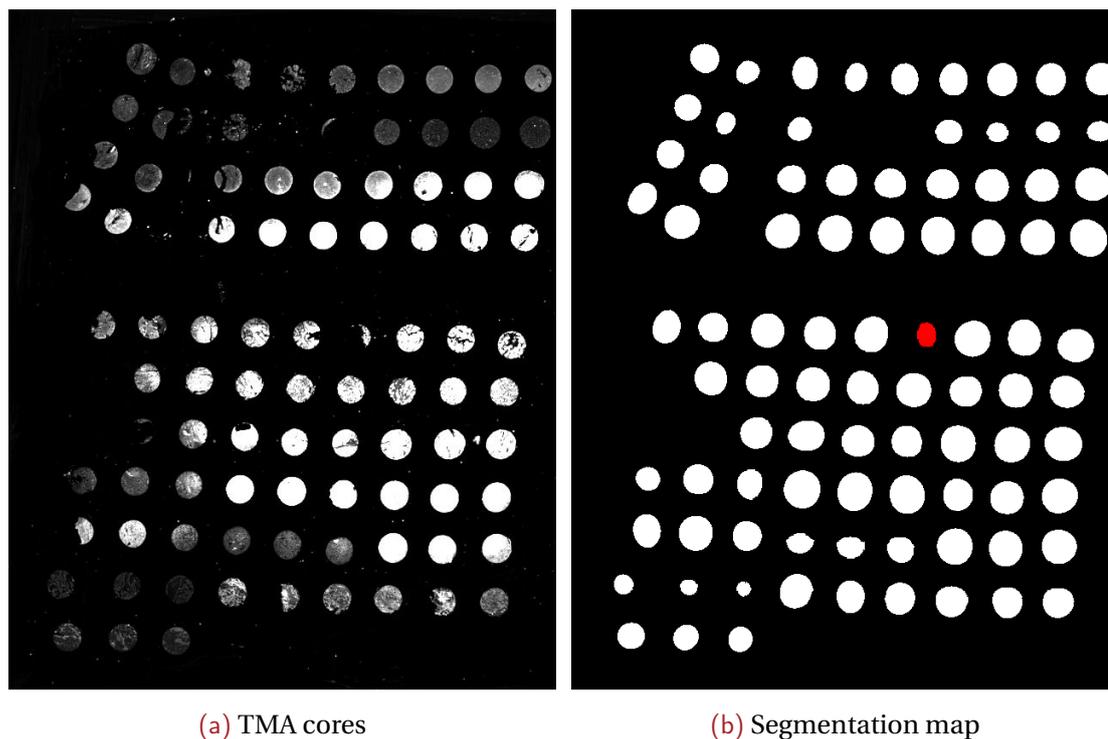
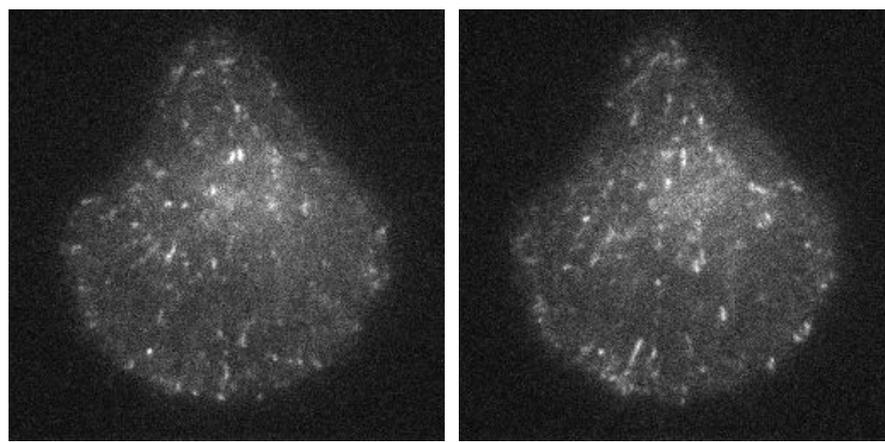


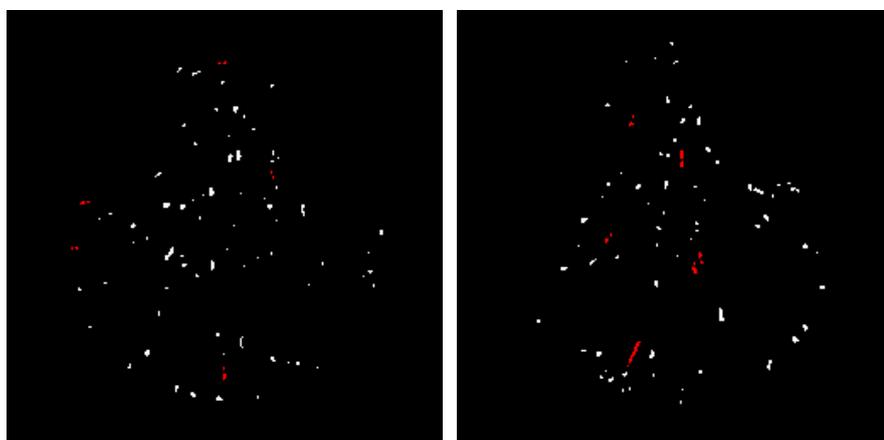
Figure 5.12 – Segmentation of cores with ATLAS. (a) Input TMA 16-bit image (courtesy of Innopsys Company). The brightest core is 76 times brighter than the darkest one. Some spurious spots are 36 times brighter than the darkest spot. (b) Segmentation results supplied by ATLAS. Only one error is made (in red), by detecting a broken core of small spatial extent.

5.4 Discussion

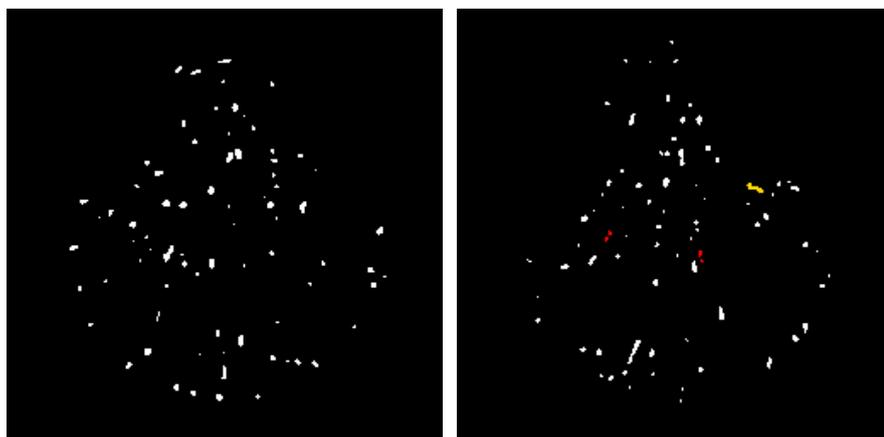
We have proposed a novel and efficient vesicle segmentation method called ATLAS which involves an automatic scale selection and a local threshold setting. It is dedicated to situations where most of the visible structures share about the same size in the image. The selected scale can be of any arbitrary precision. After determining the optimal scale, a LoG operator is applied on the images. The segmentation threshold is automatically and locally set according to a given PFA value. Overall, ATLAS outperforms state-of-the-art methods on various datasets, including a new one we have constructed and made publicly available for further comparison. Satisfactory segmentation results on several challenging real TIRFM images have been reported. We have shown that ATLAS is not sensitive to the Gaussian window size in the segmentation step. Moreover, the PFA value is a user-friendly parameter which allows the user to adapt the method to the targeted detection sensitivity according to the application needs and the further exploitation of the detection results. Thus, no specific knowledge is required on the algorithm itself, that is, the method can be used as a black box by someone non-expert in image processing. We have shown that ATLAS can be successfully applied to different kinds of images. We have also demonstrated that ATLAS



(a) Input TIRFM images



(b) Segmentation results with MS-VST



(c) Segmentation results with ATLAS

Figure 5.13 – Comparison of segmentation results on a real 300-frame TIRFM sequence presenting elongated objects. (a) Two input TIRFM images (Rab11-mCherry) out of the 300-frame sequence at time instants 93 and 300. (b) Segmentation results with MS-VST for the two frames of (a). A total of 24,481 objects are detected over the sequence. Some elongated objects of (a) are split (red) by MS-VST. (c) Segmentation results with ATLAS for the two frames of (a). A total of 24,195 objects are detected over the sequence. Few very close objects of (a) are merged (marked in yellow) by ATLAS, and the elongated objects are well recovered.

can deal with a couple of scales if needed. We will further investigate the detection and exploitation of a wider set of scales if one or two scales are not sufficient to accurately describe the structures of interest. We also plan to apply *ATLAS* to three-dimensional images.

II

Individual dynamics

6 Modeling of the vesicle fusion dynamics

RECOGNIZING dynamic protein behaviors in live cell fluorescence microscopy is of paramount importance to understand cell mechanisms. In the case of membrane traffic, cargo molecules are transferred from a donor to an acceptor compartment [Prydz et al., 2013]. During the exocytosis process, a vesicle conveys cargo molecules to the plasma membrane, and then opens to expel them from the cell. At each step, dedicated molecular platforms are acting to transport, steer and deliver selected proteins. In microscopy imaging, this sequence of processes leads to a series of different dynamics which need to be untangled in order to understand the spatiotemporal coordination of the molecular actors. Total internal reflection fluorescence microscopy (TIRFM) is particularly well suited for focusing on the late steps of exocytosis, which occur at the plasma membrane [Brown, 2006]. However, even with this modality, it is still a challenging task to classify dynamics of vesicular movements and protein diffusion.

In this part, we investigate the dynamics of three proteins associated with vesicles during the exocytosis process of vesicle recycling: Rab11, Transferrin receptor (TfR) and Langerin.

Rab11 is a so-called GTPase protein. Among others, it is required during the vesicle transport. However, Rab11 is also known to be associated with many other actors (see, e.g., [Schwartz et al., 2007, Hutagalung and Novick, 2011]), and one objective is to decipher its dynamics in order to better understand the recycling process.

TfR and Langerin are transmembrane proteins, that is, they are inserted in biological membranes, including the plasma membrane, where they are involved in several biological processes. After they have completed a process in the cell, TfR and Langerin are recycled by a recycling endosome [Uzan-Gafsou et al., 2007], which means that they are sent to another place, where they have to carry out some function. If sent to the plasma membrane, they are transported by a recycling vesicle, which eventu-

ally fuses to the plasma membrane. At this point, the transmembrane proteins diffuse in the plasma membrane and it can be reused for the following endocytosis.

To the best of our knowledge, apart from this thesis, the issue of diffusion after vesicle fusion was only addressed in [Mele et al., 2009, Burchfield et al., 2010] with a simple model and estimation method. In these papers, very restrictive hypotheses are assumed, yielding non-realistic results. Moreover, the estimation method, which we recall later, does not exploit all the available information and lacks accuracy. Therefore, we propose to go further both in terms of modeling and estimation of the vesicle fusion dynamics.

In this chapter, we present the biological processes taking place during exocytosis in Section 6.1, before and after the vesicle fusion to plasma membrane. After presenting existing diffusion models in biology in Section 6.2, dynamical models for vesicle fusion dynamics are proposed in Section 6.4 for Rab11, TfR and Langerin. Strengths and limitations of the proposed models are discussed in Section 6.5. Fusion dynamics will then be detected and estimated in Chapter 7 and Chapter 8, respectively.

6.1 Exocytosis process

6.1.1 Sequence until vesicle fusion

Let us first present very briefly the successive steps constituting the process, which are represented in Figure 6.1. The selected material being secreted is initially contained in a donor organelle. In order to leave the latter, a vesicle forms by invagination of the organelle membrane (Figure 6.1b). The material is enclosed when the nascent vesicle closes and detaches from the donor membrane.

As illustrated in Figure 6.1c, a vesicular motor then pulls the vesicle along the cytoskeleton, which allows the vesicle and its contents to move toward the plasma membrane. Let us mention that the Rab11 protein is bound to the vesicle and most probably to a molecular motor during the transport step, and stays with the vesicle for a while after transport, as explained in Section 6.1.2.

The overall process follows a docking-tethering-fusion sequence. While the vesicle approaches the plasma membrane, specific receptors in the latter, target SNAREs or t-SNAREs in short, bind with the vesicle own receptors, v-SNAREs, constituting a SNARE complex represented in Figure 6.1d. Then, the complex pulls the vesicle toward the plasma membrane until fusion.

Fusion consists in melting the vesicle and plasma membranes. Specifically, after docking at the plasma membrane, the vesicle opens while its membrane becomes part of the plasma membrane. As depicted in Figure 6.1e, the interior of the vesicle therefore becomes the exterior of the plasma membrane, and the material which was

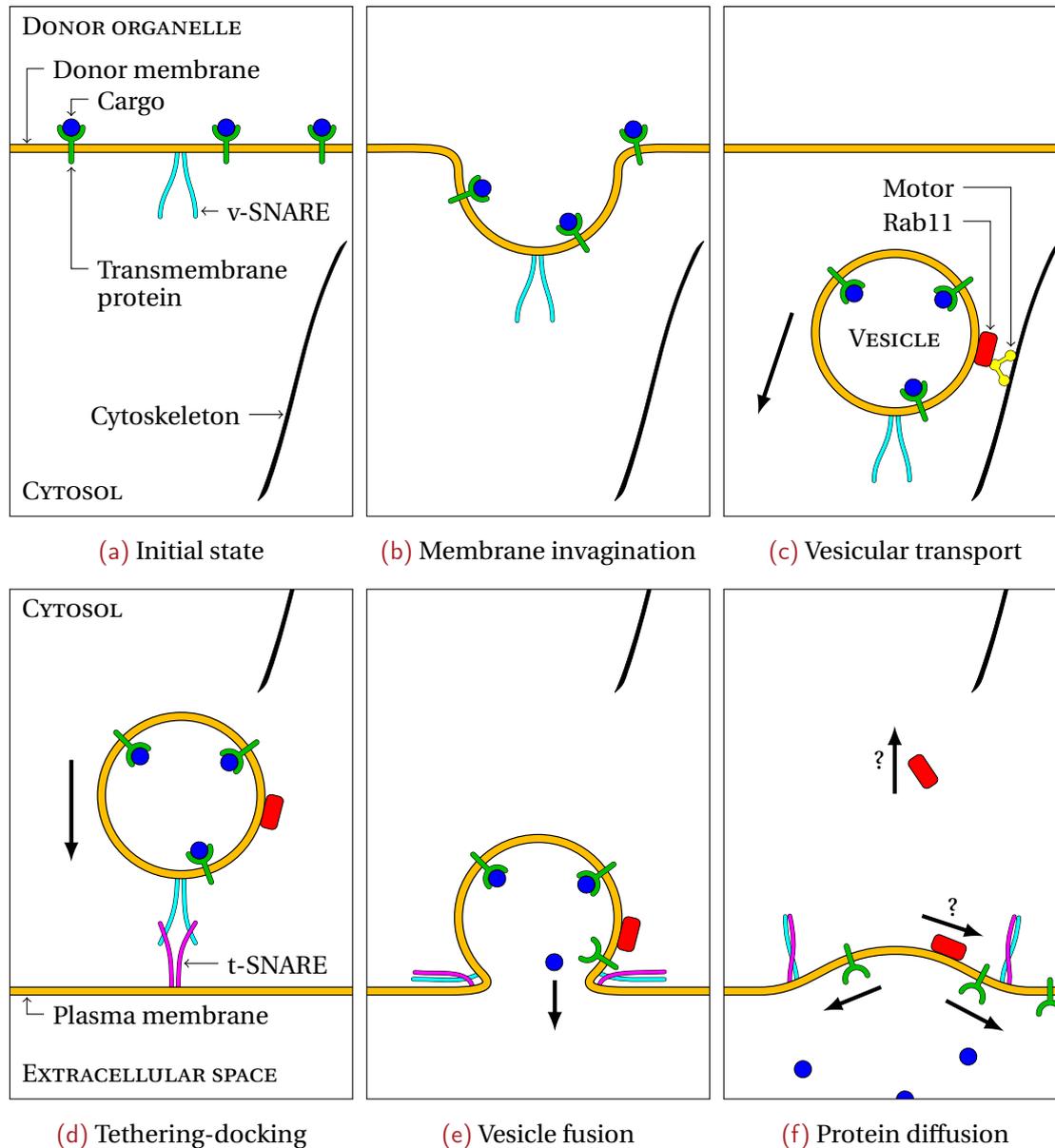


Figure 6.1 – Main steps of exocytosis

initially in the donor organelle, and then in the vesicle, is thus released outside the cell, concluding the exocytosis process. In the meantime, transmembrane proteins, which are in the vesicle membrane before fusion, are released in the plasma membrane (Figure 6.1f). The behavior of Rab11 is not easy to establish: it could be that (i) Rab11 diffuses along the plasma membrane and later dissociates from the membrane to the cytosol; or (ii) it dissociates and diffuses in the cytosol immediately after fusion. We will focus on this final step in the remaining of Part II.

As mentioned above, the behavior of TfR and Langerin on the one hand, and of Rab11 on the other hand, are very different. Specifically, TfR and Langerin are two transmembrane proteins spanning across the vesicle membrane, while Rab11 binds

at the outer side the vesicle membrane, as depicted in Figure 6.1. We shall now describe the behavior and specificity of those proteins.

6.1.2 Diffusion of Rab11 after vesicle fusion

The exact sequence of events followed by Rab11 and the associated dynamics are not well established. However, the interaction of Rab11 with many proteins during exocytosis has been demonstrated [Lindsay and McCaffrey, 2002, Hales et al., 2001, Lapierre and Goldenring, 2005, Gidon et al., 2012, Boulanger et al., 2014, Novick et al., 2006], suggesting that Rab11 is required for the successive steps of exocytosis to be completed from transport to fusion, and shows that Rab11 lies in the vicinity of the vesicle at least until fusion.

As opposed to transmembrane proteins, Rab11 can leave the membrane for the cytosol, an event named *dissociation*. While it is known that Rab11 dissociates from the plasma membrane, at some time point of the exocytosis or later, it is unclear if Rab11 dissociates before, during, or after vesicle fusion. For example, Rab11 may first move along the plasma membrane before leaving to the cytosol, that is, diffusing in two dimensions before dissociating and diffusing in three dimensions, or on the contrary, it may only diffuse in three dimensions. In both cases, deciphering the sequence of consecutive steps during and after vesicle fusion remains an open question.

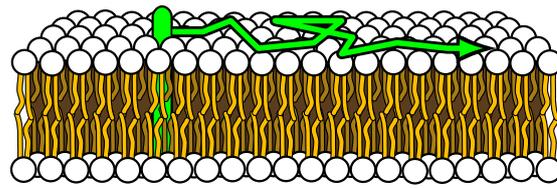
Owing to thermal agitation in the membrane or cytosol, Rab11 proteins are assumed to undergo a Brownian motion [Einstein, 1905]. This model holds if the plasma membrane and cytosol are isotropic homogeneous, and free of obstacles. Moreover the membrane is supposed to be flat in the thesis. However the hypotheses must only be valid in a small region around the vesicle fusion location, as we will rely on very local estimation methods.

Furthermore, we focus on image sequences depicting *micropatterned* cells, which means that they are “glued” on the coverslip. This ensures that the cell does not drift during the acquisition and the membrane itself is supposed to be immobile.

Proteins undergoing a Brownian motion may go away from the location where vesicle fusion occurred, and we will see that the group motion of the proteins is a *lateral diffusion* process, that is, a process by which the protein concentration tends to homogenize in the membrane or cytosol [Einstein, 1956].

To select a dynamical model for Rab11, different candidates must be proposed. We are principally interested in the dissociation process, i.e., we aim at distinguishing membrane and cytosol dynamics. For example, the following models are plausible, but others can be considered:

- When the vesicle fuses to the plasma membrane, Rab11 dissociates to diffuse in the cytosol;



© Adapted from an illustration by Mariana Ruiz Villarreal, Wikimedia Commons

Figure 6.2 – Diffusion of a transmembrane protein (in green) in the plasma membrane

- When the vesicle fuses to the plasma membrane, Rab11 starts dissociating and diffusing in the cytosol, while a decreasing portion of the Rab11 proteins diffuses along the membrane;

In the first case, we will see that a simple image model can be inferred, with only one biophysical parameter: the diffusion coefficient. In the other case, however, two concentration components superimpose in the image: (i) a component diffusing in two dimensions in the membrane; and (ii) a component diffusing in three dimensions in the cytosol. In addition, a flow model should be proposed to account for dissociation. We end up with two diffusion coefficients plus one or several parameters for the dissociation model.

6.1.3 Diffusion of TfR and Langerin after vesicle fusion

Transmembrane proteins span across the membrane, which they never leave. Therefore, their motion is constrained by the two-dimensional surface formed by the membrane they belong to. More precisely, transmembrane proteins move in the membrane by sliding between phospholipids, as illustrated in Figure 6.2. In the frame of exocytosis, TfR and Langerin either belong to the vesicle membrane or to the plasma membrane, but never dissociate to the cytosol.

Like for Rab11, the type of dynamics undergone by transmembrane proteins in the plasma membrane is assumed to be a Brownian motion [Clegg and Vaz, 1985], which holds under the aforementioned hypotheses (isotropic, homogeneous, obstacle-free membrane). More complex models have been proposed to take into account interactions between proteins and membrane structures [Sako and Kusumi, 1994, Schuss et al., 2007, Rehfeldt and Stichlmair, 2007] or to allow for rotation [Cherry, 1979, Swaminathan et al., 1997]. However, the rotational component has been shown to be negligible with respect to the lateral component [Almeida and Vaz, 1995]. For an extensive survey on the diffusion in plasma membrane, we refer the reader to [Clegg and Vaz, 1985, Almeida and Vaz, 1995].

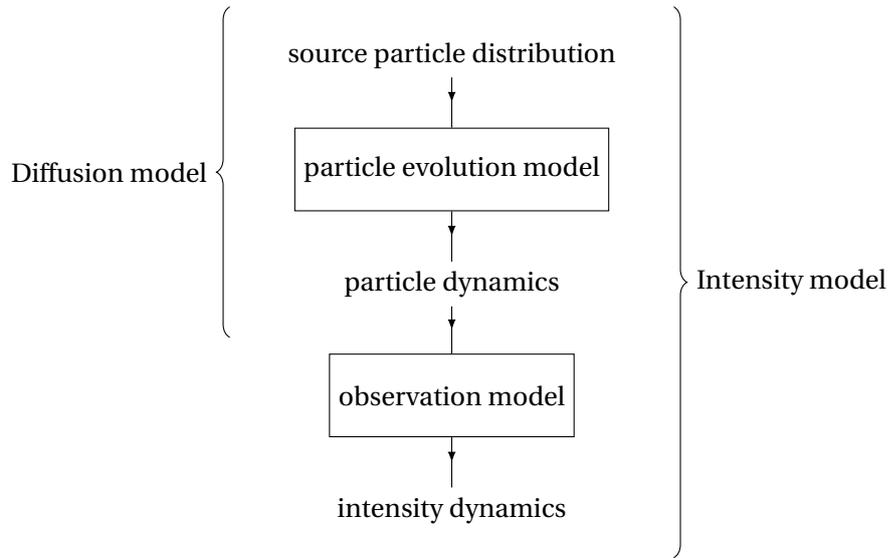


Figure 6.3 – Break-down of the intensity model

6.2 Existing diffusion models

For thirty years, numerous biological studies focused on membrane diffusion [Vaz et al., 1984, Clegg and Vaz, 1985, Jacobson et al., 1987, Almeida and Vaz, 1995, Chen et al., 2006, Guo et al., 2008, Ramadurai et al., 2009]. Different models correspond to different applications, observed objects or microscopy techniques. Before proposing a new diffusion model for vesicle fusion, let us comment three widely-used diffusion models, namely the stationary model, the point source model and the fluorescence recovery after photobleaching (FRAP) model.

Generally, the models were proposed to analyze 2D diffusion in the plasma membrane, as considered in what follows. The extension to three dimensions is generally straightforward. If this is not the case, the three-dimensional model will be explicitly given.

6.2.1 Intensity model break-down

As represented in Figure 6.3, the mathematical model of $I(p, t)$ for each point $p \in \Omega_I$ and time $t \in \mathbb{R}_+$ is fully determined by three components:

- The source particle distribution;
- The evolution model;
- The observation model.

The source distribution characterizes the way particles are spread over space before they start diffusing. The source distribution defines both the spatial distribution of the particles before they start diffusing, and the law governing their release time to

the plasma membrane or cytosol. This will be thoroughly explored in the remaining of the chapter.

The particle evolution model is the mathematical description of the motion of the proteins after fusion. As explained above, it is assumed to be Brownian. Then, resulting from Brownian motion of each individual particle, *lateral diffusion* is the dynamical model governing the evolution of the whole particle population.

As for the observation model, we have explained in Chapter 1 that it is itself subdivided into components, including different noises and the optical transfer function or point spread function (PSF). In this chapter, we will first consider a noise-free observation model to derive the intensity model.

Since the evolution and observation models are fixed once for all, we investigate in this chapter the source distribution modeling, that is the distribution of the light-emitting particles before they start diffusing.

6.2.2 From Brownian motion to lateral diffusion¹

As a fundamental stochastic process, involved for instance in fluid dynamics [Fick, 1855, Maxwell, 1867, Philibert, 2005], heat transfer [Fourier, 1822, Carslaw and Eger, 1959] or stellar dynamics [Reid and Brunthaler, 2004, Merritt, 2013], Brownian motion was extensively studied by mathematicians and physicists. While named after botanist Robert Brown, who observed such motion without a mathematical description [Brown, 1828], most mathematics of Brownian motion were formulated by Albert Einstein in the second of the *Annus mirabilis* papers [Einstein, 1905]. More recently, biophysicists and biologists introduced Brownian motion to model various dynamic structures inside the cell [Hellriegel and Gratton, 2009, Schuss, 2012], most extensively focusing on lipid and protein motion in the plasma membrane [Vaz et al., 1984, Jacobson et al., 1987, Almeida and Vaz, 1995, Saxton and Jacobson, 1997, Chen et al., 2006, Guo et al., 2008, Ramadurai et al., 2009].

Mathematically, a given particle i undergoes a Brownian motion if and only if its position over time $\mathbf{k}_i(t)$ follows a Wiener stochastic process [Wiener, 1966]. This characterization shows that the particle displacement between arbitrary times t_1 and t_2 follows a normal distribution of mean 0 and variance $2D|t_2 - t_1|$:

$$\forall t_1, t_2 \in \mathbb{R}_+, \mathbf{k}_i(t_2) - \mathbf{k}_i(t_1) \sim \mathcal{N}\left(0, 2D|t_2 - t_1|\mathbf{I}\right), \quad (6.1)$$

where \mathbf{I} denoted the 2×2 or 3×3 identity matrix depending on the diffusion space dimension.

For the problem to be well-posed, some constraints must be added. To this end, we assume that particle i is static and located at point \mathbf{k}_{0i} from $t = 0$ to a given *release*

¹Probabilistic formulation of lateral diffusion was derived with Vincent Briane during his PhD with the Serpico team at Inria.

time t_{0i} . At t_{0i} , particle i starts moving following a Wiener process:

$$\forall t \leq t_{0i}, \mathbf{k}_i(t) = \mathbf{k}_{0i}, \quad (6.2)$$

$$\forall t > t_{0i}, \mathbf{k}_i(t) \sim \mathcal{N}(\mathbf{k}_{0i}, 2D |t - t_{0i}| \mathbf{I}). \quad (6.3)$$

Local concentration

While Brownian motion describes the dynamics of individual particles, the dynamical model of a population of such particles is named *lateral diffusion* [Einstein, 1956]. To cross the bridge from Brownian motion to lateral diffusion, let us now introduce the concept of local concentration. In the vesicle, and later in the cytosol or plasma membrane, proteins of interest are numerous, so that, in TIRFM image sequences, we do not observe a single particle, but a population of N particles. Concentration is generally defined as the number of particles in a given region. Let v_p be a region around any point p . We define the *local* concentration $C(p, t)$ by reducing the region area $|v_p|$ to zero, that is, reducing the region v_p to point p :

$$C(p, t) = \lim_{|v_p| \rightarrow 0} \sum_{1 \leq i \leq N} \frac{\mathbb{1}_{v_p}(\mathbf{k}_i(t))}{|v_p|} \quad (6.4)$$

where $\mathbb{1}$ is the indicator function.

If the particles are numerous enough, independent and identically distributed, the law of large numbers yields:

$$C(p, t) \approx N f_{\mathbf{K}(t)}(p), \quad (6.5)$$

where $\mathbf{K}(t)$ is the random variable associated to the $\mathbf{k}_i(t)$'s, and $f_{\mathbf{K}(t)}$ denotes the probability density function (PDF) of $\mathbf{K}(t)$. $f_{\mathbf{K}(t)}$ can be expressed as a function of the conditional PDF $f_{\mathbf{K}(t)|\mathbf{K}_0=p_0}$, termed *transition* PDF, and the *source* PDF $f_{\mathbf{K}_0}$:

$$f_{\mathbf{K}(t)}(p) = \int_{\mathbb{R}_+} \int_{\mathbb{R}^2} f_{\mathbf{K}(t)|\mathbf{K}_0=p_0}(p) f_{\mathbf{K}_0}(p_0) f_{T_0}(t_0) \, dp_0 \, dt_0, \quad (6.6)$$

where the Wiener characterization equation (6.3) shows that $f_{\mathbf{K}(t)|\mathbf{K}_0=p_0}$ is the Gaussian PDF of mean p_0 and variance $2D |t - t_0| \mathbf{I}$. This transition PDF is specific to Brownian motion. Therefore, the whole dynamical model is only dependent on the source distribution $f_{\mathbf{K}_0}$.

Source and diffusive concentrations

The total concentration C results from the sum of the source concentration C_s and the diffusive concentration C_d :

$$C = C_s + C_d, \quad (6.7)$$

where, by definition, the particles considered for C_s do not move, while those considered for C_d undergo a Brownian motion. Specifically, C_s is defined as:

$$C_s(p, t) = \lim_{|v_p| \rightarrow 0} \sum_{1 \leq i \leq N} \frac{\mathbb{1}_{v_p}(\mathbf{k}_{0i}(t)) \mathbb{1}_{[0, T_{0i}]}(t)}{|v_p|} \quad (6.8)$$

and the law of large numbers gives:

$$C_s(p, t) \approx N f_{\mathbf{K}(t)}(p) (1 - F_{T_0}(t)), \quad (6.9)$$

where F_{T_0} is the cumulative density function (CDF) of the release time.

On the other hand, the diffusive concentration C_d is defined as:

$$C_d(p, t) \approx N f_{\mathbf{K}(t)}(p) F_{T_0}(t). \quad (6.10)$$

Diffusion equation

Now that local concentration has been introduced, let us present the lateral diffusion model which governs evolution over time of the diffusive concentration.

If we assume that t_{0i} is identical for all the particles, we merely get, with $t_0 = t_{0i}$:

$$C(p, t) = \begin{cases} C_s(p, t) & \text{for } 0 \leq t \leq t_0, \\ C_d(p, t) & \text{for } t > t_0. \end{cases} \quad (6.11)$$

In this case, the Fick's second law gives the evolution over time and space of the local concentration as a function of the diffusion coefficient D [Fick, 1855]:

$$\frac{\partial C_d}{\partial t} = D \Delta C_d, \quad (6.12)$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ (in three dimensions) denotes the Laplace operator. Based on the law of large numbers, equation (6.12) is the macroscopic counterpart of the single particle description of Brownian motion (6.1) in the sense that it characterizes the evolution of a population of such particles observed at a higher scale.

On the other hand, in the general case where t_{0i} is different for all particles, as in the proposed "small-extent source with exponential decay release" (SSED) model, a *flow* is introduced from C_s to C_d :

$$\frac{\partial C_d}{\partial t} = D \Delta C_d - \frac{\partial C_s}{\partial t}. \quad (6.13)$$

To our knowledge, we are introducing this flow in the study of membrane dynamics.

Fick's second law (6.12) can be solved by Fourier analysis, which yields the following closed form Green's function Φ [Fourier, 1822]:

$$\forall t > t_0, \forall p \in \mathbb{R}^2, \Phi(p, t) = \frac{1}{4\pi D (t - t_0)} \exp\left(-\frac{\|p - p_0\|_2^2}{4D (t - t_0)}\right). \quad (6.14)$$

Equation (6.14) can also be interpreted from the stochastic point of view as reflecting the probability of finding particles at position p and time t , if they undergo a Brownian motion of diffusion coefficient D and are initially concentrated at p_0 . Actually, instead of relying on the Fick's second law equation (6.12), the Green's function can also be viewed as a transition PDF:

$$\forall t > t_0, \forall p \in \mathbb{R}^2, f_{\mathcal{K}(t)|\mathcal{K}_0=p_0}(p) = \frac{1}{4\pi D(t-t_0)} \exp\left(-\frac{\|p-p_0\|_2^2}{4D(t-t_0)}\right). \quad (6.15)$$

This shows the equivalence of the microscopic, stochastic process and the macroscopic, deterministic partial differential equation (PDE).

6.2.3 Stationary model

In the stationary model [Sergeev, 2004], particles are initially spread with a uniform distribution, and start moving at t_0 :

$$\forall t \in \mathbb{R}_+, \forall p \in \mathbb{R}^2, C_s(p, t) = \begin{cases} C_0 & \text{for } 0 \leq t \leq t_0 \\ 0 & \text{for } t > t_0 \end{cases} \quad (6.16)$$

$$\text{and } \forall p \in \mathbb{R}^2, C_d(p, t_0) = C_0. \quad (6.17)$$

The Fick's second law then gives the temporal evolution of C_d :

$$\forall t > t_0, \forall p \in \mathbb{R}^2, C_d(p, t) = C_0, \quad (6.18)$$

which yields:

$$\forall t \in \mathbb{R}_+, \forall p \in \mathbb{R}^2, C(p, t) = C_0. \quad (6.19)$$

In turn, the stationary model is the multi-particle extension of the Brownian motion model, where particles themselves undergo a Brownian motion, but the local concentration C is expected not to vary in space and time, thus the stationary moniker.

When the particle number N is high, and the law of large numbers holds, the intensity I is thus expected to be constant, as illustrated in Figure 6.4a.

In contrast, when the particle density is low, intensity is not constant. Indeed, in this case, the microscope PSF gives the particles a Gaussian-like profile in the image, resulting in a sum of Gaussian spots, whose center positions are sampled from the uniform distribution (Figure 6.4b). Parameters of the model are C_0 , the diffusion coefficient D and the radius σ_{PSF} of the PSF.

As presented in Chapter 8, many papers dealt with the problem of estimating the diffusion coefficient D under the stationary model, but it can only be used for low enough particle densities; otherwise, the intensity model does not even depend on D .

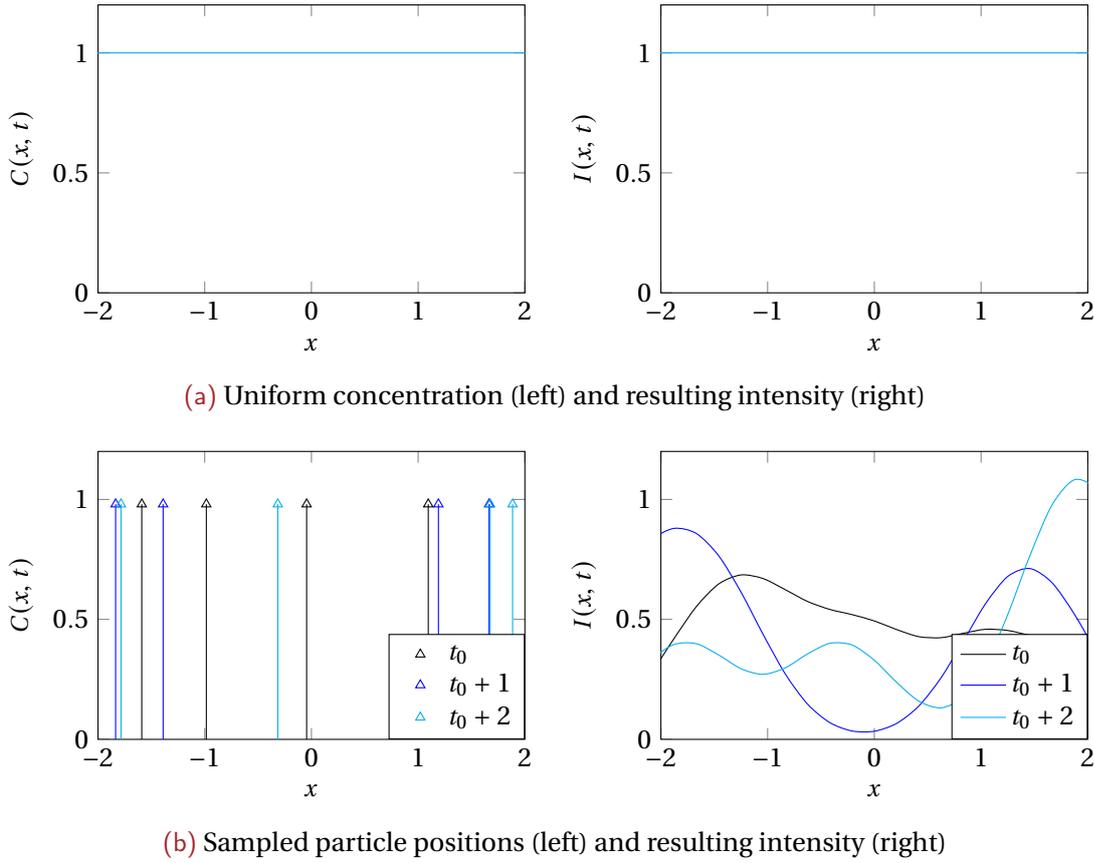


Figure 6.4 – Stationary model at three time steps for $C_0 = 1$, $D = 1$, $\sigma_{\text{PSF}} = 1$. Arrows represent the Dirac distribution.

6.2.4 Point source model

In the point source model, we assume that all the particles are initially concentrated at p_0 and that they all start diffusing at time t_0 , that is, the source distribution is proportional to a spatiotemporal Dirac distribution:

$$\forall t \in \mathbb{R}_+, \forall p \in \mathbb{R}^2, C_s(p, t) = C_0 \delta(p - p_0) \delta(t - t_0). \quad (6.20)$$

The Green's function (6.14) is the solution of the Fick's second law for a space-time Dirac source distribution. Then, by linearity of the Fick's second law, the concentration dynamics is merely obtained by multiplying the Green's function by C_0 :

$$\forall t > t_0, \forall p \in \mathbb{R}^2, C(p, t) = \frac{C_0}{4\pi D(t - t_0)} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0)}\right). \quad (6.21)$$

Then, to derive the intensity model, we need to incorporate the observation model, reduced to the PSF and gain of the microscope. In two dimensions, that is, for membrane diffusion, the PSF is modeled as a two-dimensional Gaussian function of variance σ_{PSF}^2 [Small and Stahlheber, 2014], so I is obtained by convolving the con-

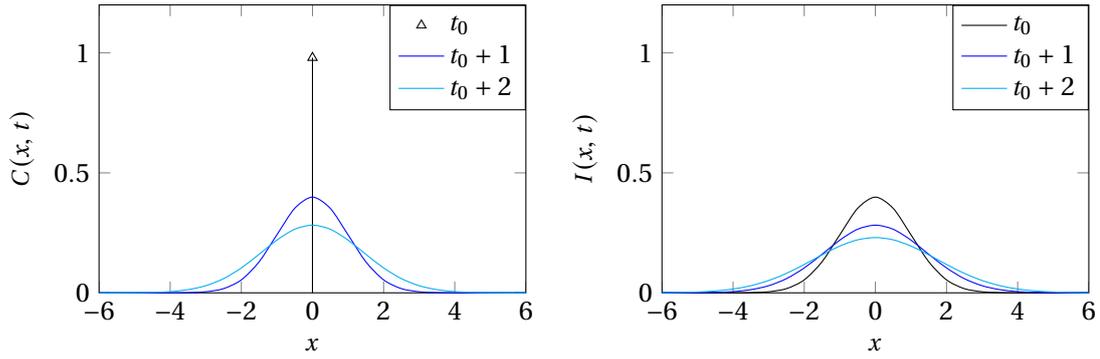


Figure 6.5 – Point source model at three time steps for $C_0 = 1$, $D = 1$, $p_0 = 0$, $\sigma_{\text{PSF}} = 1$. Arrow represents the Dirac distribution.

centration (6.14) with a Gaussian kernel of variance σ_{PSF}^2 , which gives:

$$\forall t \geq t_0, \forall p \in \mathbb{R}^2, I(p, t) \propto \frac{C_0}{4\pi D(t - t_0) + 2\pi\sigma_{\text{PSF}}^2} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0) + 2\sigma_{\text{PSF}}^2}\right). \quad (6.22)$$

For the sake of simplicity, we introduce constant A_0 such that:

$$\forall t \geq t_0, \forall p \in \mathbb{R}^2, I(p, t) = \frac{A_0}{2D(t - t_0) + \sigma_{\text{PSF}}^2} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0) + 2\sigma_{\text{PSF}}^2}\right). \quad (6.23)$$

The source distribution and expected concentration of the point source model are plotted in Figure 6.5.

Since it is very practical for computation and estimation purposes, the point source model was used in [Mele et al., 2009, Burchfield et al., 2010] to estimate the dynamics of vesicle fusion. To this end, two supplementary hypotheses were assumed:

H1 The vesicle is pointwise;

H2 The whole transmembrane protein material is released instantaneously at t_0 .

However, we will see that these hypotheses are not always justified, which will lead us to relax them and propose a more elaborated model.

To our knowledge, the extension of [Mele et al., 2009, Burchfield et al., 2010] to the 3D diffusion has not been done yet, but it will be proposed in Section 6.3.3 as an ingredient of the Rab11 dynamics modeling.

6.2.5 FRAP model²

In FRAP experiments, transmembrane proteins are supposed to be uniformly distributed with an average concentration denoted C_0 . At time t_0 , a laser beam locally kills fluorescence by photobleaching [Axelrod et al., 1976, Cherry, 1979, Kapitza and Jacobson, 1986, Seiffert and Oppermann, 2005]. Bleached transmembrane proteins then stop emitting photons, so that they do not appear anymore in C or I .

²FRAP theory and tools were presented to us by François Waharte at UMR 144.

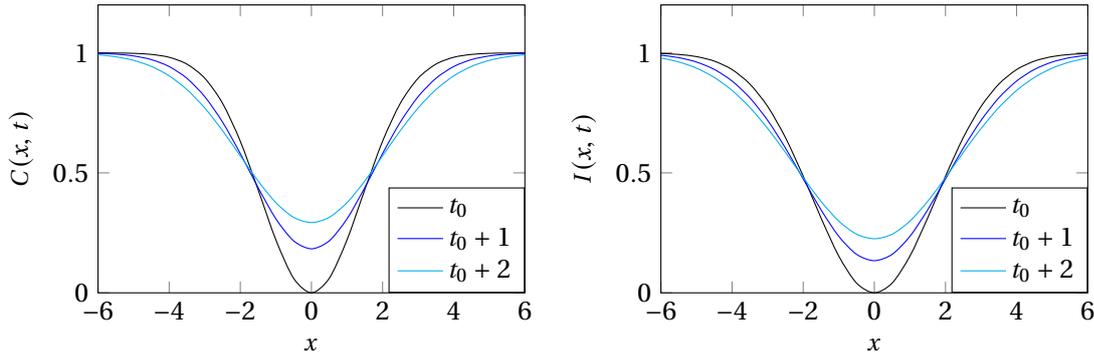


Figure 6.6 – FRAP model at three time steps for $C_0 = 1$, $D = 1$, $p_{\text{beam}} = 0$, $\sigma_{\text{beam}} = 2$

To ease computation, the beam profile is often approximated by a Gaussian function of center p_{beam} and radius σ_{beam} [Seiffert and Oppermann, 2005]. After t_0 , fluorescent (resp. bleached) particles, which undergo a Brownian motion, repopulate the bleached (resp. fluorescent) region, until both bleached and fluorescent particles are uniformly spread. Then, the intensity becomes uniform again. Some extended models were proposed, e.g., to account for uniformly distributed static particles. We refer the reader to [Carrero et al., 2003] for more details on extended FRAP models.

Owing to the superposition principle, the solution of the Fick's second law (6.12) for an arbitrary source profile $s : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ is obtained by convolving the Green's function with the profile. In FRAP experiments, the laser beam is generally assumed to have a Gaussian profile, resulting in a Gaussian “hole” fluorescence profile presented in Figure 6.6:

$$s(p) = \frac{1}{2\pi\sigma_{\text{beam}}^2} - \frac{1}{2\pi\sigma_{\text{beam}}^2} \exp\left(-\frac{\|p - p_{\text{beam}}\|_2^2}{2\sigma_{\text{beam}}^2}\right). \quad (6.24)$$

The intensity model is straightforwardly deduced by convolution with the PSF, as the convolution of two Gaussian functions is the Gaussian function whose variance is the sum of the original variances:

$$I(p, t) = \frac{A_0}{\sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2} - \frac{A_0}{2D(t - t_0) + \sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0) + 2\sigma_{\text{PSF}}^2 + 2\sigma_{\text{beam}}^2}\right). \quad (6.25)$$

Since this solution is analogous to the point source model solution, any method for estimating the point source model can be used in FRAP experiments.

A three-dimensional extension is generally proposed in the form of the so-called diffusion-dissociation model. As the name implies, particles leave the membrane and disappear out of reach for the TIRFM evanescent wave. Dissociation is generally introduced in the model by multiplying equation (6.25) with a spatially constant, expo-

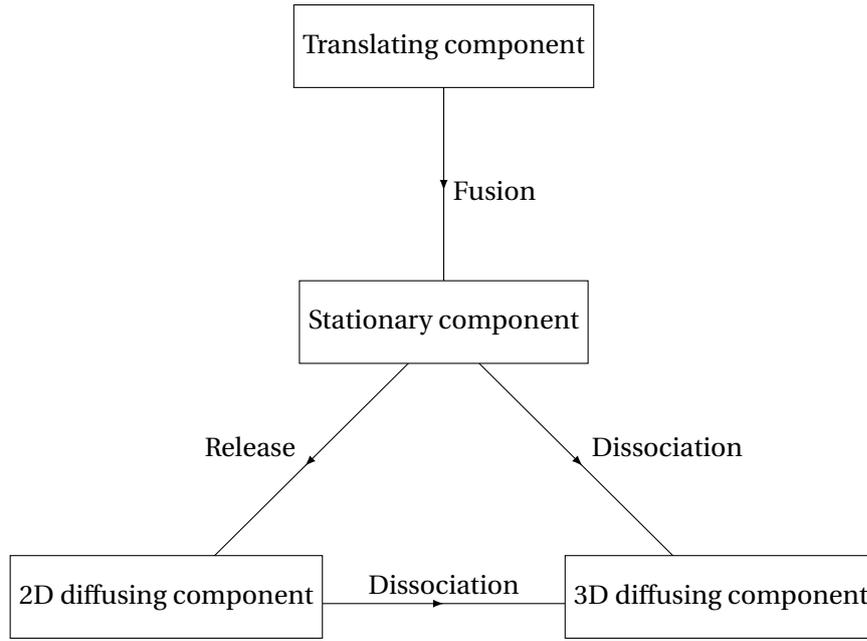


Figure 6.7 – Particle evolution meta-model for the Rab11 dynamics

nential decay function:

$$\begin{aligned}
 I(p, t) &= \exp\left(-\frac{t}{\tau}\right) \\
 &\times \left[\frac{A_0}{\sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2} - \frac{A_0}{2D(t - t_0) + \sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0) + 2\sigma_{\text{PSF}}^2 + 2\sigma_{\text{beam}}^2}\right) \right].
 \end{aligned}
 \tag{6.26}$$

6.3 Proposed fusion models for Rab11

While the sequence of events of Rab11 is not totally established yet, we propose a *meta-model* to encompass different hypothetical models. It is illustrated in Figure 6.7. Numerous models compatible with current biological knowledge can be instantiated from it by defining flows between a stationary state (which accounts for the proteins remaining in the vesicle after fusion) and three dynamical states, namely *translation* (from transport to docking), *membrane diffusion* and *cytosol diffusion*. We consider the following flows:

- Fusion specifies the beginning of the diffusion(s);
- Release characterizes the flow between static concentration and concentration diffusing in the plasma membrane;
- Dissociation is the process by which Rab11 leaves the membrane, regardless of its previous state; as explained in [Carrero et al., 2003, Michelman-Ribeiro et al., 2009, Im et al., 2013], it is modeled as an exponential decay of the observed concentration.

Let us detail the dynamical models for translation and the two diffusions.

6.3.1 Vesicle motion before fusion

We assume that the frame rate (10 frames/s in the processed sequence) is sufficient to neglect rotation and deformation of the vesicle. Hence, the vesicle is supposed to undergo a translational motion between two successive time points t and $t + 1$. Let $V_i^\star(t)$ denote the spatial support at time t of such a vesicle. By definition of the translation model, all the pixels of the connected component $V_i^\star(t)$ formed by the vesicle at time t share the same displacement w_i :

$$\forall p \in V_i^\star(t), w_i(p, t) = w_i(t), \quad (6.27)$$

where $w_i(p, t)$ denotes the displacement of $p \in V_i^\star(t)$ between t and $t + 1$.

6.3.2 Two-dimensional diffusion along the plasma membrane

In [Mele et al., 2009], the point source model was used to model the diffusion in the plasma membrane after vesicle fusion. In contrast, in order to take into account the size of the vesicle, we consider a finite spatial support (e.g., a spherical vesicle) of diameter $2r$, that is, the spatial source profile $s : \mathbb{R}^2 \rightarrow \mathbb{R}$ is null outside a disc $\mathcal{B}_r \subset \mathbb{R}^2$ of radius r . Without loss of generality, let us assume that $\int_{\mathbb{R}^2} s = 1$, and that \mathcal{B}_r is centered in $(0, 0)$. Let I_r (resp. I_δ) be the solution of the Fick's second law for the source of profile s (resp. for the point source case).

Proposition 4. The difference between I_r and I_δ is small when r is small. More precisely:

$$|I_r - I_\delta| = \mathcal{O}\left(\frac{r}{\sigma_{\text{PSF}}}\right), \quad (6.28)$$

by introducing the Landau \mathcal{O} notation.

Proof. Let us up-bound the absolute difference between I_r and I_δ :

$$I_r(p, t) - I_\delta(p, t) = \int_{\mathbb{R}^2} \frac{A_0}{2\pi\sigma^2(t)} \exp\left(-\frac{\|p - q\|_2^2}{2\sigma^2(t)}\right) s(q) \, dq - \frac{A_0}{2\pi\sigma^2(t)} \exp\left(-\frac{\|p\|_2^2}{2\sigma^2(t)}\right),$$

with $\sigma^2(t) = 2D(t - t_0 + \sigma_{\text{PSF}}^2)$.

Since $\int_{\mathbb{R}^2} s = 1$, we get:

$$I_r(p, t) - I_\delta(p, t) = \frac{A_0}{2\pi\sigma^2(t)} \int_{\mathbb{R}^2} \left[\exp\left(-\frac{\|p - q\|_2^2}{2\sigma^2(t)}\right) - \exp\left(-\frac{\|p\|_2^2}{2\sigma^2(t)}\right) \right] s(q) \, dq.$$

And since $s(q) = 0$ when $q \notin \mathcal{B}_r$, the integration is made over \mathcal{B}_r :

$$I_r(p, t) - I_\delta(p, t) = \frac{A_0}{2\pi\sigma^2(t)} \int_{\mathcal{B}_r} \left[\exp\left(-\frac{\|p - q\|_2^2}{2\sigma^2(t)}\right) - \exp\left(-\frac{\|p\|_2^2}{2\sigma^2(t)}\right) \right] s(q) \, dq.$$

We will now exhibit an upper bound of the bracket. To this end, let us introduce the following auxiliary function:

$$\begin{aligned} h: \mathbb{R}_+ &\rightarrow \mathbb{R}_+ \\ x &\mapsto \exp\left(-\frac{x^2}{2\sigma^2(t)}\right). \end{aligned}$$

From the mean value theorem, we get:

$$\begin{aligned} \left| h(\|p - q\|_2) - h(\|p\|_2) \right| &\leq \left| \|p - q\|_2 - \|p\|_2 \right| \max|h'| \\ &\leq \|q\|_2 \max|h'|, \end{aligned}$$

with

$$\begin{aligned} \left| h'(\|p\|_2) \right| &= \frac{\|p\|_2}{\sigma^2(t)} \exp\left(-\frac{\|p\|_2^2}{2\sigma^2(t)}\right) \\ &< \frac{1}{\sigma(t)}. \end{aligned}$$

Thus:

$$\begin{aligned} |I_r(p, t) - I_\delta(p, t)| &= \frac{A_0}{2\pi\sigma^2(t)} \int_{\mathcal{B}_r} \left| h(\|p - q\|_2) - h(\|q\|_2) \right| s(q) \, dq \\ &< \frac{A_0}{2\pi\sigma^2(t)} \int_{\mathcal{B}_r} \frac{\|q\|_2}{\sigma(t)} s(q) \, dq \\ &< \frac{A_0}{2\pi\sigma^2(t)} \frac{r}{\sigma(t)} \int_{\mathcal{B}_r} s(q) \, dq \\ &< \frac{A_0}{2\pi\sigma^2(t)} \frac{r}{\sigma(t)} \\ &< \frac{A_0}{2\pi\sigma^2(t)} \frac{r}{\sigma_{\text{PSF}}}. \end{aligned}$$

□

This shows that, whatever the vesicle profile, if its diameter is small enough with respect to the PSF width, the point source model remains accurate. Otherwise, the model becomes accurate only after a given time interval, when $\sigma(t)$ is high enough with respect to the vesicle diameter.

6.3.3 Three dimensional diffusion in the cytosol for Rab11

Proteins diffusing in the cytosol also follow the Fick's second law (6.12), but the Laplacian operator now acts in a 3D space, that is:

$$\frac{\partial C_d}{\partial t} = D \Delta C_d = D \left(\frac{\partial^2 C_d}{\partial x^2} + \frac{\partial^2 C_d}{\partial y^2} + \frac{\partial^2 C_d}{\partial z^2} \right). \quad (6.29)$$

This yields a Green's function analogous to (6.14):

$$\Phi(p, t) = \frac{1}{(4\pi D(t - t_0))^{3/2}} \exp\left(-\frac{\|p - p_0\|_2^2}{4D(t - t_0)}\right), \quad (6.30)$$

where $p = (x, y, z)^T$.

However, as fluorophores are present at $z > 0$, the PSF cannot be reduced to a simple two-dimensional Gaussian profile, as done for diffusion along the plasma membrane. Let us derive the intensity model for the point source in three dimensions observed with TIRFM.

The exponential decay of the excitation field of TIRFM must be integrated, which gives, with d the penetration depth:

$$\begin{aligned} C_d(x, y, t) &= \int_0^\infty C_d(x, y, z, t) \exp\left(-\frac{z}{d}\right) dz \\ &= \frac{C_0}{(4\pi D(t - t_0))^{3/2}} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{4D(t - t_0)}\right) \int_0^\infty \exp\left(-\frac{z^2}{4D(t - t_0)} - \frac{z}{d}\right) dz \\ &= \frac{C_0}{(4\pi D(t - t_0))^{3/2}} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{4D(t - t_0)}\right) \\ &\quad \times \sqrt{\pi D(t - t_0)} \exp\left(\frac{D(t - t_0)}{d^2}\right) \left[1 - \operatorname{erf}\left(\frac{\sqrt{D(t - t_0)}}{d}\right)\right] \\ &= \frac{C_0}{8\pi D(t - t_0)} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{4D(t - t_0)}\right) \exp\left(\frac{D(t - t_0)}{d^2}\right) \left[1 - \operatorname{erf}\left(\frac{\sqrt{D(t - t_0)}}{d}\right)\right]. \end{aligned} \quad (6.31)$$

This equation could be integrated numerically, so that parameters could be estimated, but an approximation will exhibit the issue of studying the 3D+t model in 2D+t image sequences. Let us rely on the following exponential approximation of the error function [Chiani et al., 2003]:

$$\operatorname{erf}(\sqrt{x}) \approx 1 - \frac{1}{2} \exp(-x). \quad (6.32)$$

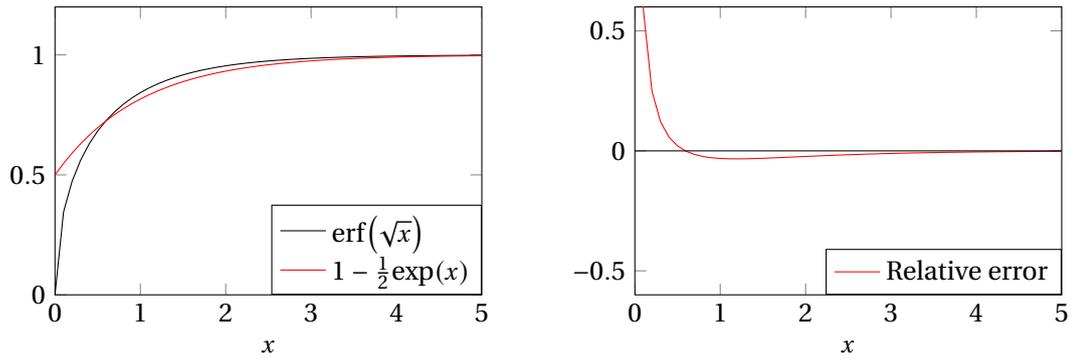


Figure 6.8 – Approximation of the error function

As plotted in Figure 6.8, the approximation is precise when $x \geq 1$. For $x = D(t - t_0)/d^2$, this condition corresponds to $t - t_0 \geq d^2/D$, which rapidly holds after t_0 ; orders of magnitude of d and D are 1 px and 1 px²/frame so that $t - t_0 \geq d^2/D$ at $t = t_0 + 1$ already. In this case, we get:

$$\operatorname{erf}\left(\frac{\sqrt{D(t - t_0)}}{d}\right) \approx 1 - \frac{1}{2} \exp\left(-\frac{D(t - t_0)}{d^2}\right). \quad (6.33)$$

This yields:

$$C_d(p, t) \approx \frac{C_0}{16\pi D(t - t_0)} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{4D(t - t_0)}\right), \quad (6.34)$$

that is, we get the same solution as for the point source problem in two dimensions, up to a factor. While this means that we can apply the same estimation procedure to get the diffusion coefficient of Rab11 for this model, this raises an issue. Actually, we are not precisely interested in the diffusion coefficient estimation, but in the selection of a dynamical model. Therefore, as two- and three-dimension diffusion models have very close forms, 2D TIRFM it not sufficient to reliably discriminate the models, even for the simplest two- and three-dimensional point source models. Naturally, the situation would be even more complex and intractable if we would have also considered a continuous release and dissociation models.

For now, we have no 3D+time TIRFM images with a frame rate sufficient to study sub-second dynamics. However, we are confident that the technology will be ready in a few months [Boulangier et al., 2014], thus allowing to conduct the study of Rab11 dynamical behavior in the near future.

6.4 Proposed SSED model for TfR and Langerin

In order to take into account the non-instantaneous release of the proteins at fusion, we propose a new fusion model for TfR and Langerin.

Table 6.1 – Characterization of the diffusion models

Model	Source distribution		Parameters*
	Spatial distribution	Temporal distribution	
Stationary	Uniform	Dirac	$C_0, D, \sigma_{\text{PSF}}$
Point source	Dirac	Dirac	$C_0, D, p_0, \sigma_{\text{PSF}}$
FRAP	Gaussian hole	Dirac	$C_0, D, p_{\text{beam}}, \sigma_{\text{beam}}$
SSED	Arbitrary (with small support)	Exponential decay	$C_0, \tau, D, p_0, \sigma_{\text{PSF}}$

* C_0 : initial local concentration; D : diffusion coefficient; σ_{PSF} : radius of the PSF; p_{beam} : center of the laser beam; σ_{beam} : radius of the laser beam; p_0 : position of the vesicle; τ : residence time.

The proposed “small-extent source with exponential decay release” (SSED) model is briefly compared to the aforementioned existing models in Table 6.1. The SSED model consists in relaxing both the pointwise source (H1) and instantaneous release (H2) hypotheses of the point source model.

6.4.1 Continuous release of the concentration

In Figure 6.9, we compare the evolution of concentration over time in true and simulated diffusion image sequences. Precisely, fusion events in sequences showing TfR and Langerin are represented by the means of kymographs. A kymograph gives the evolution in time of a given image line, by appending its successive profiles. Namely, in Figure 6.9b, the line $x = 136$ is used to display two TfR fusion events. The evolution of line $x = 161$ in the second sequence shows another fusion event, for Langerin this time. Then, simulated kymographs are presented, both for the point source and SSED models. Apart from quantitative aspects, we observe that the respective intensity profiles are very different, especially long after vesicle fusion. In particular, the central intensity peak observed after t_0 in real image sequences (mainly for Langerin) cannot be obtained with the point source model.

As we have shown that the pointwise source hypothesis is indeed valid, the other hypothesis, the instantaneous release, must be wrong. Therefore, instead of assuming that all the particles start diffusing at t_0 , we now consider a continuous release of the particles, where each particle is expected to stay at p_0 during a certain amount of time τ called residence time:

$$\frac{\partial C_s}{\partial t} = -\frac{1}{\tau} C_s. \quad (6.35)$$

This yields an exponential decay of the source concentration:

$$\forall t \geq t_0, \forall p \in \mathbb{R}^2, C_s(p, t) = C_0 \delta(p - p_0) \exp\left(-\frac{t - t_0}{\tau}\right). \quad (6.36)$$

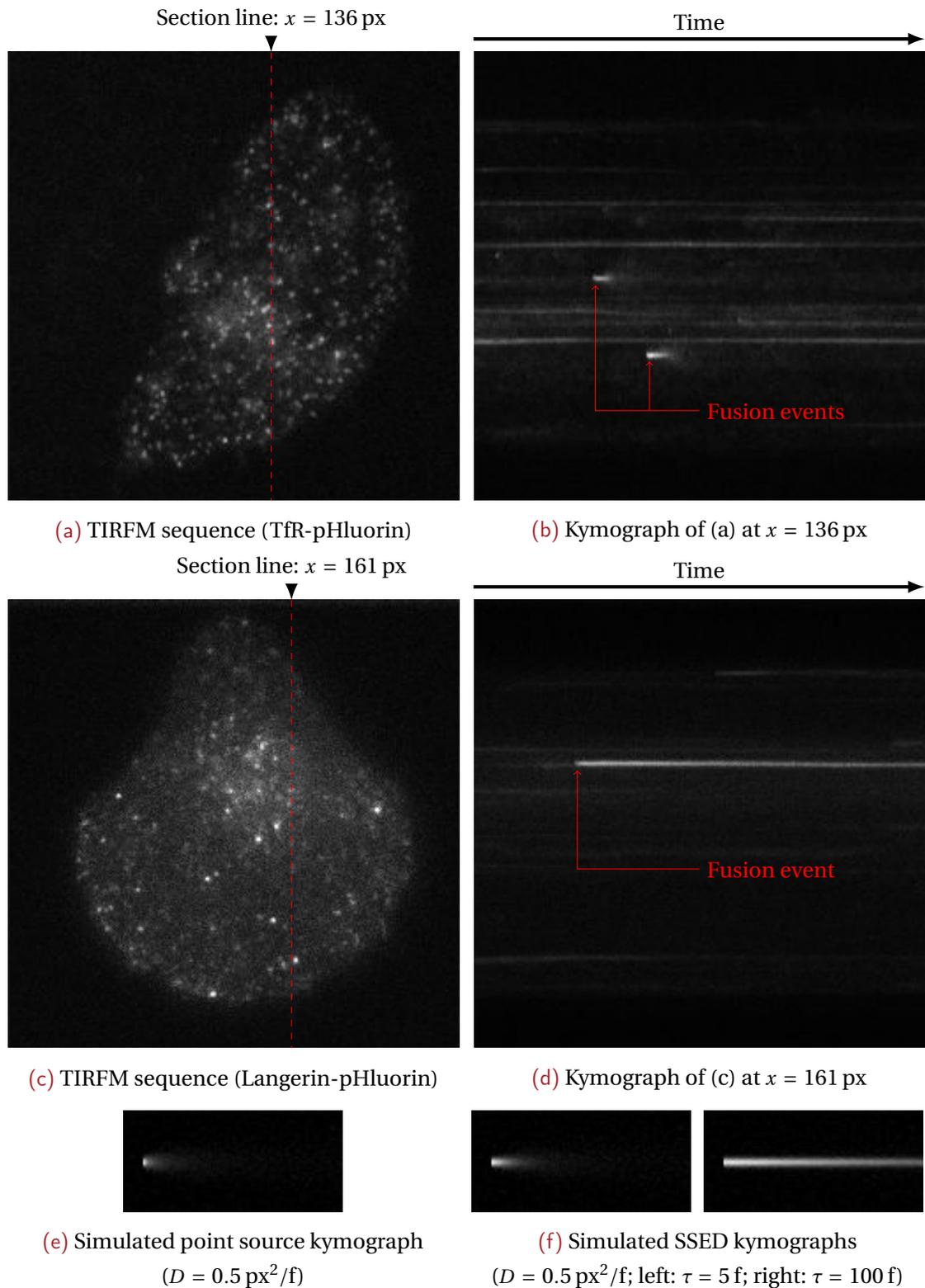


Figure 6.9 – Comparison of real image sequences with simulations of the point source and SSED models

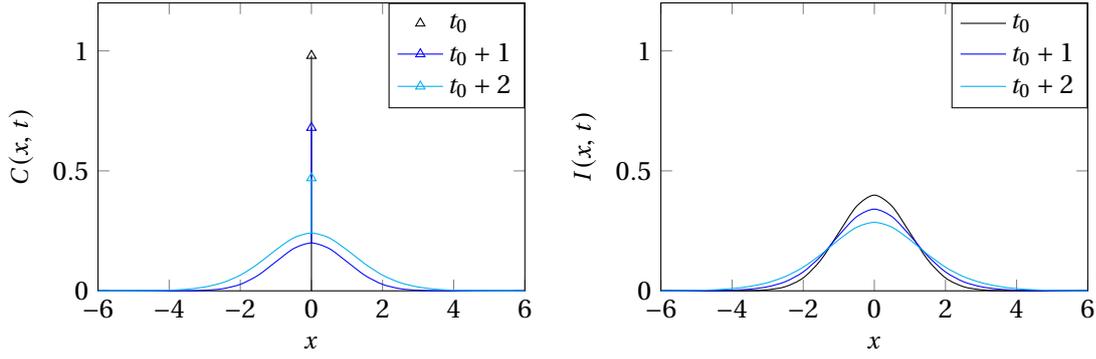


Figure 6.10 – SSED model at three time steps for $C_0 = 1$; $\tau = 2$. $D = 1$; $p_0 = 0$; $\sigma_{\text{PSF}} = 1$. Arrows represent the Dirac distribution.

This model is illustrated in Figure 6.10. Different underlying dynamics can yield such an exponential decay release, e.g., a narrow escape [Schuss et al., 2007, Singer et al., 2008, Schuss, 2012], corresponding to the release of proteins through a small aperture, or dissociation-like process [Carrero et al., 2003, Michelman-Ribeiro et al., 2009, Im et al., 2013]. Other hypotheses could yield other continuous release models, but the exponential decay has the advantage of being easy to handle from an estimation point of view. Actually, another (maybe even simpler) model is the constant flow model, where a constant amount of concentration is released at each time instant. However, this linear decrease has to stop when all the material is released, resulting in non differentiable functions. We are aware that the exponential decay is a simple model, but it is both a plausible and practical choice.

The proteins which leave the source compartment diffuse in the plasma membrane. Thus, the flow between both compartments acts as a source in the diffusion equation (6.13), which yields:

$$\begin{aligned} \forall t > t_0, \forall p \in \mathbb{R}^2, \frac{\partial C_d}{\partial t}(p, t) - D \Delta C_d(p, t) &= -\frac{\partial C_s}{\partial t}(p, t) \\ &= \frac{C_0}{\tau} \delta(p - p_0) \exp\left(-\frac{t - t_0}{\tau}\right). \end{aligned} \quad (6.37)$$

Still owing to the superposition principle, the solution of this equation is obtained by convolving the Green's function (6.14) with the source function (6.36):

$$\forall t > t_0, \forall p \in \mathbb{R}^2, C_d(p, t) = \int_{t_0}^t \frac{C_0/\tau}{4\pi D(t-u)} \exp\left(-\frac{u - t_0}{\tau} - \frac{\|p - p_0\|_2^2}{4D(t-u)}\right) du. \quad (6.38)$$

Finally, the observed concentration is the sum of both components:

$$C = C_s + C_d, \quad (6.39)$$

and the intensity model, illustrated in Figure 6.10, is obtained after convolution by the PSF:

$$\forall t \geq t_0, \forall p \in \mathbb{R}^2, I(p, t) = \frac{A_0/\tau}{\sigma_{\text{PSF}}^2} \exp\left(-\frac{t}{\tau} - \frac{\|p - p_0\|_2^2}{2\sigma_{\text{PSF}}^2}\right) + \int_{t_0}^t \frac{A_0}{2D(t-u) + \sigma_{\text{PSF}}^2} \exp\left(-\frac{u-t_0}{\tau} - \frac{\|p-p_0\|_2^2}{4D(t-u) + 2\sigma_{\text{PSF}}^2}\right) du. \quad (6.40)$$

6.4.2 Arbitrary spatial source distribution

Interestingly, the proof of Section 6.4.2 holds for the SSED model, so that a pointwise source is not required either for equation (6.40) to be valid. The vesicle only needs to be small enough with no assumption on its shape.

This way, we have relaxed both the pointwise and instantaneous release hypotheses of the point source model used in [Mele et al., 2009].

6.5 Discussion

In this chapter, we have constructed new diffusion models for vesicle fusion to plasma membrane. Both for Rab11 and transmembrane proteins, we have shown that a small vesicle model is equivalent to a pointwise model, which relaxes this generally assumed hypothesis, and enables the use of the point source model with fewer constraints than in the literature.

Unfortunately, we have also shown that 2D and 3D diffusion models observed by the means of 2D TIRFM are barely distinguishable. Therefore, future experiments should rely on the nascent 3D TIRFM [Boulanger et al., 2014], which should provide superior data quality to allow for better dynamics classification. When such image sequences will show up, estimation methods will be extended to the depth dimension. In the meanwhile, the diffusion coefficient can still be estimated in 2D TIRFM image sequences (see Chapter 8).

As for the continuous release introduced for transmembrane proteins, we have proposed an exponential decay model. While more complex by one parameter than the point source model, the so-built SSED better corresponds to observed dynamics, and can still be estimated, as proposed in Chapter 8.

Further extensions to the SSED model could target the so-called kiss-and-run dynamics [Rizzoli and Jahn, 2007, Miklavc et al.,], assuming that the vesicle only partially empties before reforming and leaving the membrane.

7

Detection of diffusion dynamics

BEFORE ESTIMATING the parameters of the models representing the membrane fusion, the events of interest, which we name *fusion events*, must be detected. Fusion events are defined by the space-time point at which a vesicle fusion starts, that is, the location and time at which the protein of interest is released to the plasma membrane or cytosol.

Depending on the protein observed in the image sequence, there are three cases:

- Rab11 is tagged with mCherry;
- TfR or Langerin is tagged with pHluorin;
- Dual View is used to observe both Rab11-mCherry and TfR-pHluorin or both Rab11-mCherry and Langerin-pHluorin.

In the first case, the vesicle is observed while moving before diffusing, while in the second case, owing to the pH-sensitivity of pHluorin, the fluorescence intensity is very low before fusion. This imposes different strategies to detect fusion events.

In the case of Rab11-mCherry, we propose a likelihood test to distinguish translation from diffusion dynamics. For transmembrane proteins tagged with pHluorin, fusion events start with a spot appearance, so a simple spot detection can be applied to the temporal frame difference sequence. This method can also be employed to study events in Rab11 images when Dual View with TfR or Langerin is available.

The chapter is composed of two sections. The detection of fusion events in image sequences depicting only Rab11 is addressed in Section 7.1. In Section 7.2, the properties of pHluorin are exploited to detect the events in image sequences depicting transmembrane proteins.

7.1 Detection based on Rab11 dynamics classification

In order to detect Rab11 fusion events, we propose a method based on a likelihood test to select the most appropriate dynamical model representing the vesicle dynamics in each frame. Specifically, we test the *translation* model, valid for the transport to docking steps, against the *diffusion* model, valid from fusion onward. For each detected vesicle, the transition frame, that is the first frame when the diffusion model is selected, merely corresponds to t_0 . The fusion location p_0 corresponds to the vesicle location at t_0 .

Let us assume for now that the parameters for the translation and diffusion models have been estimated – this will be presented in the next chapter. Then, we will merely select the most likely model. We have thus developed a three-step detection method. First, spots are detected relying on the spot detector proposed in Part I. Secondly, static spots are discarded owing to a statistical argument; only moving and diffusing spots then remain. Dynamics are finally classified as either translation or diffusion. The fusion event begins when the spot dynamics changes from the former to the latter.

The whole classification process only relies on three consecutive frames used to detect time-varying vesicles and evaluate temporal gradients.

7.1.1 Segmentation of time-varying vesicles

We first need to extract the vesicles that apparently change over time, or time-varying vesicles, in the TIRFM image sequence, that is, the vesicles which are moving or diffusing. The segmentation of the time-varying vesicles combines the detection of spatial spots and of significant temporal changes.

Spatial spot segmentation

We naturally use our adaptive thresholding of LoG images with auto-selected scale (ATLAS) detector (see Part I) which provides a binary vesicle presence map. In each frame $I(t)$ of the sequence, we end up with a set of connected components $\mathcal{V}(t) = \{V_i(t)\}$.

Detection of temporal changes

In order to handle the temporal dimension of the segmentation of the time-varying vesicles, we examine the evolution of the segmentation map over time by using the change detection algorithm introduced in [Boulangier et al., 2010a]. At each space-time location (p, t) in the TIRFM image sequence, we consider a 2D block $\eta(p, t)$ centered at that location and we compute the minimum of the intensity similarity dis-

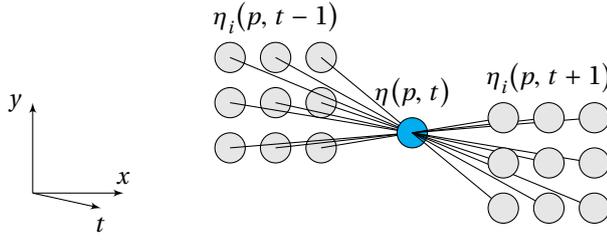


Figure 7.1 – Illustration of the temporal block neighborhood

tance (sum of square difference) of this block to neighboring blocks taken in the previous frame and in the next frame. Nine neighboring blocks $\eta_i(p, t - 1)$ are considered in the previous frame, and nine neighboring blocks $\eta_i(p, t + 1)$ are considered in the next frame, as represented in Figure 7.1.

The minimum distance d_{\min} for all the pixels in the image sequence is assumed to follow a generalized extreme value (GEV) distribution [Boulanger et al., 2010a], defined by its CDF as:

$$F(d_{\min}(p, t), \alpha, \beta, \kappa) = \exp \left[- \left(1 - \kappa \frac{d_{\min}(p, t) - \beta}{\alpha} \right)^{1/\kappa} \right], \quad (7.1)$$

where α , β and κ are respectively the width, location and shape parameters of the GEV distribution. The parameters are estimated by using a mixed L-moments/maximum likelihood method [Boulanger et al., 2010a].

We set a probability of false alarm (PFA) in order to derive a threshold (quantile of the GEV distribution) able to detect the significant temporal changes in the image between two time points. The resulting significant change map is denoted $E(t)$ (see [Boulanger et al., 2010a] for more details).

Map of time-varying vesicles

Finally, to obtain the set $\mathcal{V}^*(t) = \{V_i^*(t)\}$ of time-varying vesicles, we simply discard every vesicle of $\mathcal{V}(t)$ whose intersection with $E(t)$ is empty:

$$\mathcal{V}^*(t) = \{V_i(t) \in \mathcal{V}(t) \mid \exists q \in V_i(t), E(p, t) = 1\}. \quad (7.2)$$

The spatiotemporal segmentation workflow is illustrated in Figure 7.2. To save computation time, we evaluate the complete similarity distance map, estimate the associated GEV parameters, and deduce the change detection threshold from the first three frames of the sequence only. Afterwards, we evaluate the similarity distance only for points belonging to the vesicles $V_i(t)$ in the subsequent images. Moreover, since the threshold has already been evaluated, not all distances must be evaluated for each vesicle.

Finally, only translating and diffusing vesicles belong to $\mathcal{V}^*(t)$.

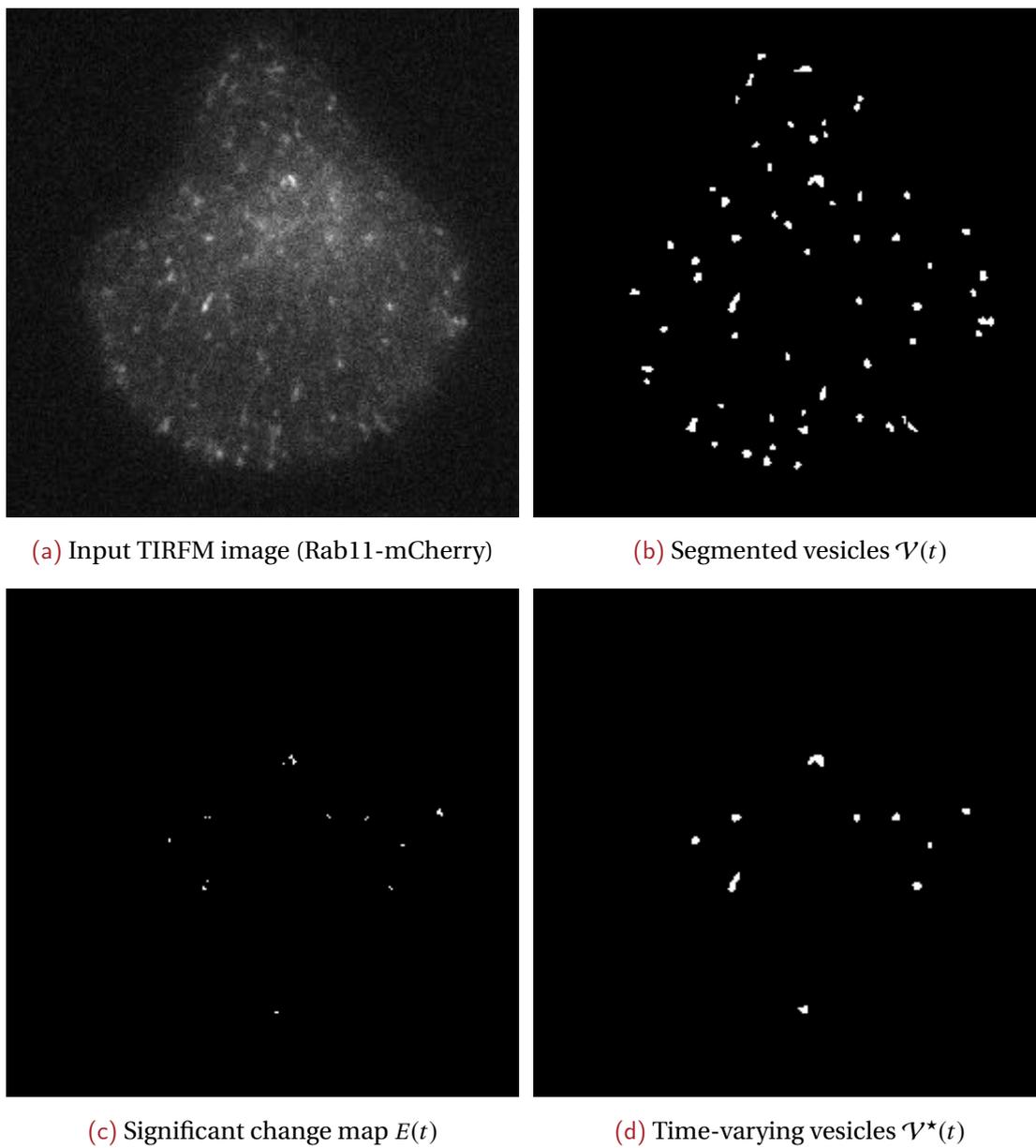


Figure 7.2 – Segmentation of the time varying vesicles

7.1.2 Classification of dynamical events

As explained above, we have to decide for a given vesicle $V_i^*(t)$ at time point t , whether it undergoes a translational motion or a diffusion. The two competing hypotheses are the following ones:

- H_0 : the vesicle $V_i^*(t)$ is undergoing a translation with displacement \mathbf{w}_i ,
- H_1 : the vesicle $V_i^*(t)$ is undergoing a diffusion with coefficient D_i .

We consider statistical image models to derive the corresponding likelihood functions.

To assess local motion, we use the displaced frame difference (DFD) given by:

$$\varepsilon_0(p, \mathbf{w}_i, t) = I(p + \mathbf{w}_i(p, t), t + 1) - I(p, t). \quad (7.3)$$

For diffusion modeling, we can straightforwardly infer the random variable ε_1 from the Fick's second law, which yields:

$$\varepsilon_1(p, t) = I_t(p, t) - D_i \Delta I(p, t), \quad (7.4)$$

where I_t denotes the temporal gradient of I .

We assume that both variables ε_0 and ε_1 are independent and identically distributed, and follow a zero-mean Gaussian distribution. As a consequence, the likelihood functions are defined as the product of Gaussian densities over the vesicle connected component $V_i^*(t)$. Then, the decision relies on the following log-likelihood ratio test:

$$\log \left(\frac{\psi_1(V_i^*(t), D_i)}{\psi_0(V_i^*(t), \mathbf{w}_i)} \right) > \tau. \quad (7.5)$$

If the ratio is greater than τ , H_1 is selected and H_0 otherwise.

7.1.3 Experimental results

Synthetic image sequences¹

In order to validate our approach, simple but realistic image sequences were simulated. The microtubules and cortical Actin are modeled with a set of 3D random splines. While one extremity of each microtubule is connected to a centrosome, the other one is located near the plasma membrane. The Actin lies mostly in the first 100 nm slice of the cell. Finally, the vesicles are modeled by a set of points whose evolution is governed by interactions with the cytoskeleton. The action of molecular motors such as myosin and kinesin is represented by an elastic force between the particle and the second nearest point to the cytoskeleton. The module of the force

¹Synthetic sequences were generated by Jérôme Boulanger at UMR 144.

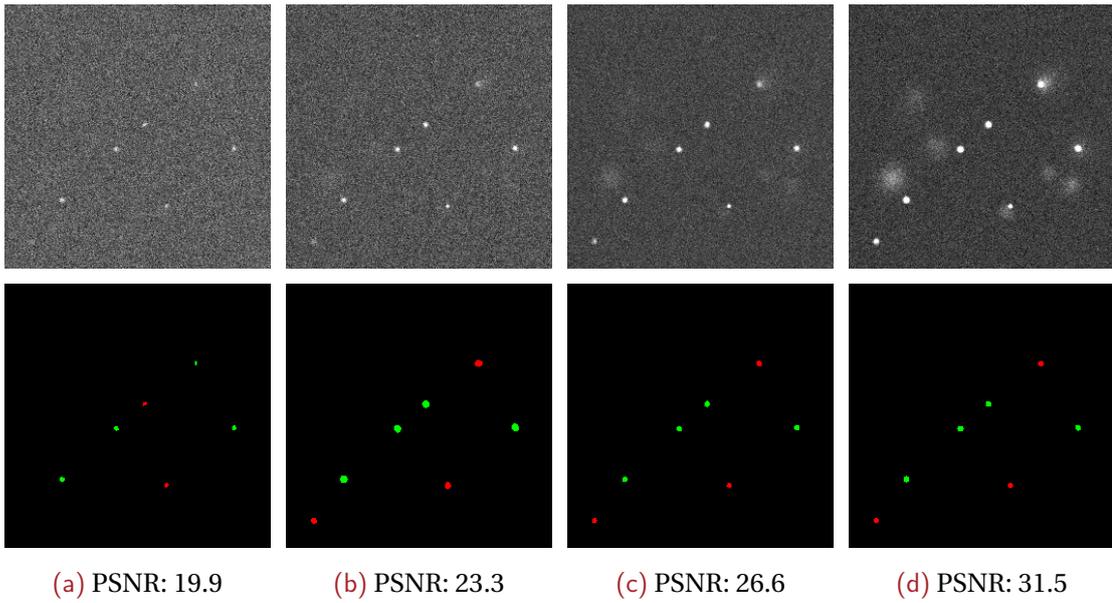


Figure 7.3 – Classification results for several synthetic sequences. Top row: First frame of the sequence for different PSNRs. Wide spots are vesicles that have already diffused. Bottom row: Vesicles classified as diffusing are in red, vesicles classified as translating are in green. In all these examples, classification is performed with no error, except for the lowest noise level (a), where two vesicles are misclassified.

depends on the type of interaction between the particle and the cytoskeleton to account for the diversity of the molecular motors involved during vesicle traffic. The spatial confinement is handled with a dedicated potential and we consider an additional viscosity term in the evolution model. Let us note that only the vesicles evolve while the cytoskeleton remains still, which is true for this time scale. Finally, when a vesicle reaches the plasma membrane, a finite difference scheme is used to model the diffusion process by the Fick's second law.

To get a realistic amount of noise, we have estimated the peak signal-to-noise ratio (PSNR) over 12 real TIRFM sequences depicting M10 cells where Rab11 is tagged with mCherry. The PSNR of these TIRFM sequences ranges from 28.3 to 31.5. According to the above described model, we have generated several synthetic sequences corrupted with different levels of Poisson-Gaussian noise. The estimated PSNR of the synthetic sequences ranges from 19.9 to 31.5 (we have used the same procedure to estimate the noise of the real sequences and fix the noise level in the synthetic sequences).

For all the sequences, we have used the same segmentation parameters: estimation windows of radius 50 px, Gaussian p-value for the segmentation of 1×10^{-3} and GEV p-value for the event detection of 5×10^{-3} . Each sequence contains about 1500 vesicles distributed in 300 frames of size 256×256 pixels. The first frame of a few sequences and the associated classification map are depicted in Figure 7.3 for various noise levels.

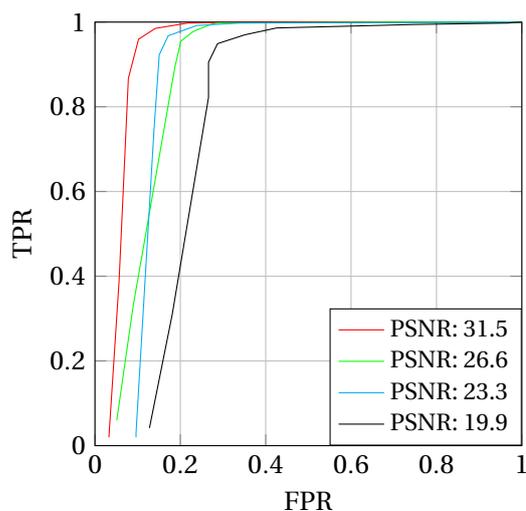


Figure 7.4 – FROC curves of the translation class for different noise levels. The curves are obtained by varying τ . Since only two classes are considered, the FROC curve for the diffusion class can be deduced from this one.

As reported in Figure 7.4, the classification results are improved if the PSNR values are high. However, for a low PSNR as 23.3 (that is much lower than that of the real TIRFM image sequences), true positive rate (TPR) is already higher than 0.99 for a false positive rate (FPR) lower than 0.2.

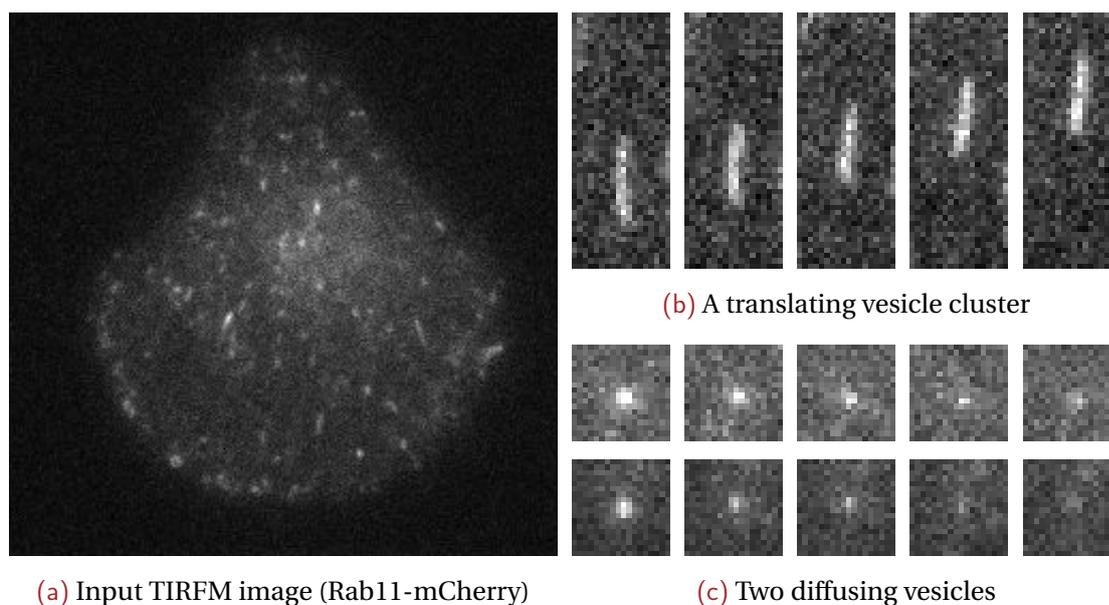
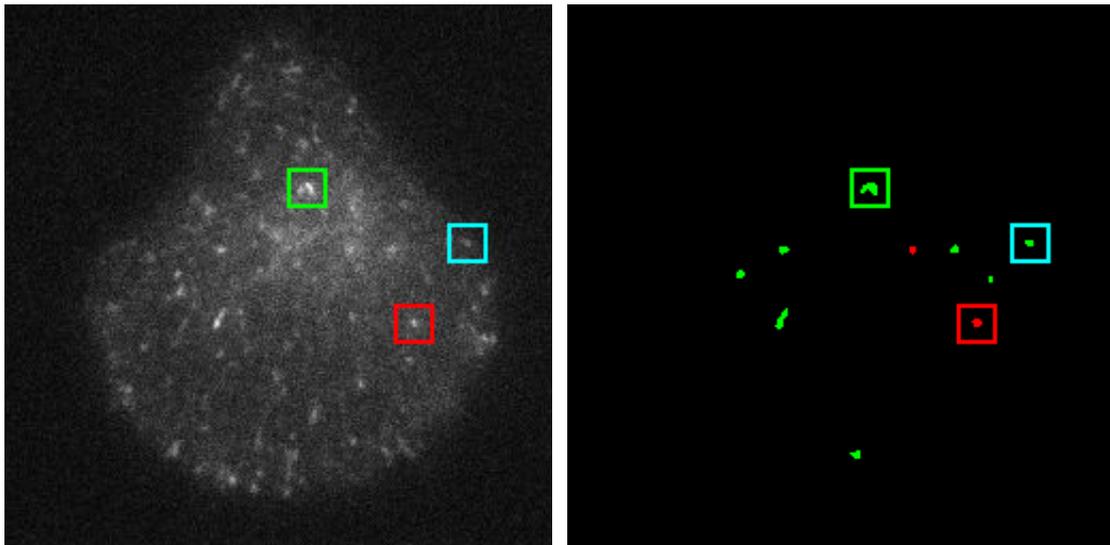


Figure 7.5 – Examples of translating and diffusing vesicles in a real TIRFM sequence depicting a M10 cell. (a) First frame of the sequence. (b) Patch in successive frames depicting an elongated translating vesicle cluster. (c) Patches in successive frames depicting two diffusing vesicles.



(a) Input TIRFM image (Rab11-mCherry)

(b) Classification results

Figure 7.6 – Classification results for a real TIRFM sequence, whose estimated PSNR is 28.6. Results are displayed for a representative frame. The only classification error – framed in cyan – is a diffusion classified as translation. However, this vesicle has a very low intensity and changes its shape while diffusing. Two vesicles framed in green are detected as a single connected component. The diffusing vesicle framed in red corresponds to the upper one of Figure 7.5c.

Real TIRFM image sequences

We have applied the method to real TIRFM sequences depicting M10 cells transfected with fluorescently labeled Rab11. The sequences are composed of 300 frames of size 256×256 pixels. The lowest estimated PSNR is 28.6, but the background is far more complicated than those of the synthetic sequences. Examples of translating and diffusing vesicles in one of these sequences are displayed in Figure 7.5.

Segmentation results for this sequence are shown in Figure 7.2, and the classification obtained for this frame is displayed and commented in Figure 7.6. Classification errors mostly correspond to unstable behavior over the time for the diffusion case, based on the instantaneous classification approach.

Globally, the results on these challenging sequences are promising, but we will see that a more efficient and much simpler approach allows to reliably detect fusion events relying on Dual View system and pH-sensitivity of pHluorin.

7.2 Event detection using fluorophore pH-sensitivity

When TfR and Langerin are tagged with pHluorin, fluorescence increases with pH. This tag is actually employed because the pH of the vesicle is about 5.5, while that of extracellular medium is expected to be 7. Acidic pH inside the vesicle before t_0 leads to very low pHluorin photon emission. When the vesicle fuses to the plasma

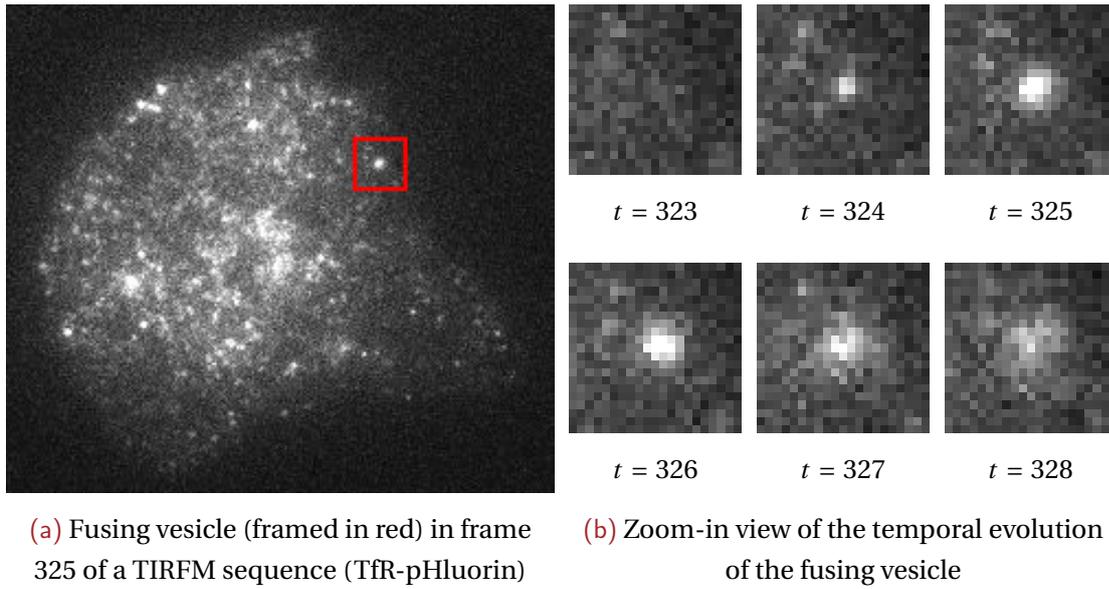


Figure 7.7 – Temporal evolution of the region around a fusion event

membrane, the pHluorin gets exposed to the neutral extracellular medium, so that the fluorescence suddenly increases, as shown in Figure 7.7. This means that the pHluorin emits very few photons while inside the vesicle, but the intensity dramatically increases when the vesicle opens, i.e., precisely at fusion time and location. Therefore, starting of fusion coincides with the sudden appearance of a spot in the image sequence.

7.2.1 Fusion event detection

Hence, we aim at detecting localized rapid fluorescence increases of intensity in I . To this end, we rely on the temporal backward difference $I_{\delta t}$ defined as:

$$\forall p \in \Omega_I, \forall t \in \mathbb{N}^*, I_{\delta t}(p, t) = I(p, t) - I(p, t - 1). \quad (7.6)$$

A fusion event $e_i = (p_{0i}, t_{0i})$ corresponds to a bright spot centered at p_{0i} in the map $I_{\delta t}(t_{0i})$. To detect events, we rely on an adapted version of the spot detection method ATLAS (see Part I).

First, the scale of the vesicles s^* is automatically selected in a multiscale representation of the images $I(t)$. We use the first ten frames of the input sequence I , as it contains more spots than ten frames of $I_{\delta t}$, and ten frames is more than enough to observe hundreds of spots.

In fact, scale selection in I (instead of $I_{\delta t}$) is appropriate, because numerous static spots are present, aside the few fusing vesicles. Therefore, the scale of the static spots is selected, which also correspond to the scale of diffusing spots, at the time they appear and start diffusing. The found scale corresponds to the scale of the spots we aim at detecting in $I_{\delta t}$.

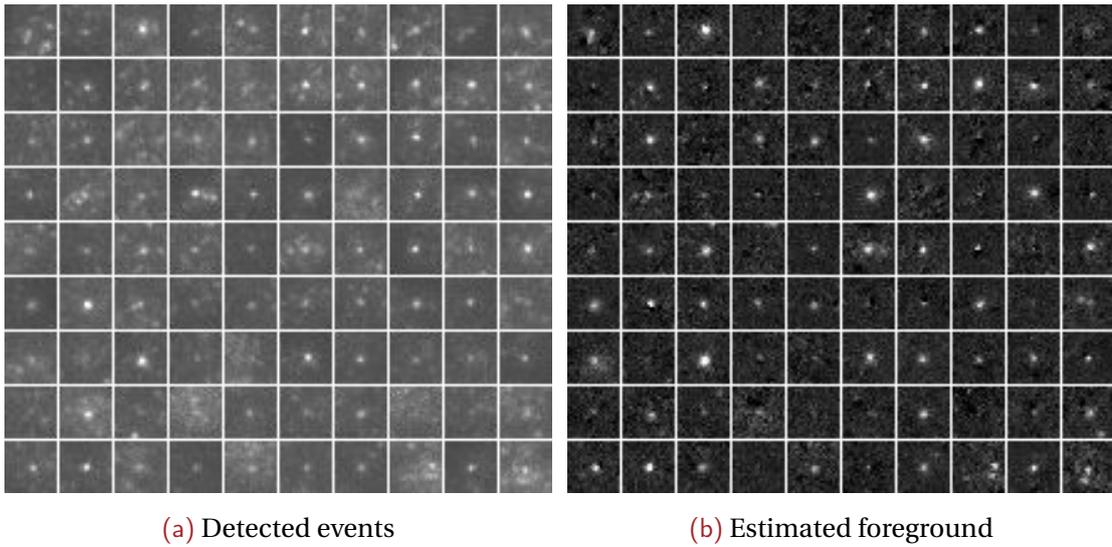


Figure 7.8 – Background subtraction in the neighborhood of membrane events detected in the TIRFM sequence of Figure 7.7

Secondly, appearing spots related to a fusion event are detected by thresholding the Laplacian of Gaussian (LoG) of scale s^* of every $I_{\delta t}(t)$, $t \in \mathbb{N}^*$. As described in Part I, the threshold automatically adapts to local LoG statistics estimated in a sliding Gaussian window, whose size is not critical. Its radius is set to 60 px, which is a trade-off on background structure sizes in the processed images. The detection threshold is inferred pointwise from a PFA fixed to 1×10^{-6} . We end up with a set of N fusion events $\{\widehat{e}_i\}_{i=1..N}$ detected over the image sequence.

Let ν_i be a spatiotemporal neighborhood of e_i and let I_i denote the restriction of I to ν_i . From observation of real TIRFM images and quantitative evaluation on simulations, we have chosen a spatial neighborhood of 21×21 pixels centered at \widehat{p}_{0i} , and a temporal range of 5 frames from \widehat{t}_{0i} to $\widehat{t}_{0i} + 4$ for the point source model, and 20 frames from \widehat{t}_{0i} to $\widehat{t}_{0i} + 19$ for the SSED model. Background structures and static vesicles may exist close to the fusing spot, and then constitute outliers of the Gaussian spot model. To get rid of them, we need to subtract the image background as follows. The background intensity $b_i(q)$ at image point $q \in \nu_i$ is estimated as the median of the intensity values at q over the twenty frames acquired before \widehat{t}_{0i} :

$$\widehat{b}_i(q) = \text{median}_{t_{0i}-20 \leq t < t_{0i}} I_i(q, t). \quad (7.7)$$

Then, subtracting $\widehat{b}_i(q)$ for $q \in \nu_i$ merely gives us the estimated foreground subsequence $\widehat{z}_i = I_i - \widehat{b}_i$. As shown in Figure 7.8, most background structures are removed.

7.2.2 Prior selection of diffusion processes under the point source model

Nevertheless, we observe that detected fusion events are not always followed by a point source diffusion process. Actually, some detections correspond to spots which

apparently only fade out without spreading. As shown in the previous chapter (Figure 6.9f), this case possibly corresponds to events with long residence time under the SSED model. However, to estimate the point source model parameters, we need to retain only spots which really satisfy it.

While classical goodness-of-fit tests can be employed after the model parameters have been estimated, we propose a method to select diffusing events *before* estimating all the parameters, which reduces computation time, as the diffusion coefficient will be estimated only for actual diffusion processes.

Let us focus on a given event e_i detected at \widehat{p}_{0i} in frame \widehat{t}_{0i} . If e_i corresponds to a point source diffusion process, equation (6.23) holds in z_i . Let us subscript with i the parameters associated to the diffusion process in v_i .

- t_{0i} is the fusion event beginning;
- p_{0i} is the fusion event location;
- A_i is the initial spot amplitude;
- D_i is the diffusion coefficient.

According to equation (6.14), the protein concentration is Gaussian-distributed in every frame after t_0 . Let $A_i(t)$ and $\sigma_i^2(t)$ respectively denote the amplitude and variance of the observed Gaussian spot at time instant t . They are supposed to vary during the diffusion process. We will use \widehat{t}_{0i} as an approximation of t_{0i} – as shown below, \widehat{t}_{0i} could even be chosen arbitrarily higher than t_{0i} to estimate D_i . Then, $A_i(\widehat{t}_{0i})$ and $\sigma_i^2(\widehat{t}_{0i})$ are the “initial” spot amplitude and variance, and equation (6.23) yields:

$$z_i(p, t) = A_i(t) \exp\left(-\frac{\|p - p_{0i}\|_2^2}{2\sigma_i^2(t)}\right), \quad (7.8)$$

$$\text{where } A_i(t) = \frac{A_{0i}}{2D_i(t - t_{0i}) + \sigma_{0i}^2} \quad (7.9)$$

$$\text{and } \sigma_i^2(t) = 2D_i(t - t_{0i}) + \sigma_{0i}^2. \quad (7.10)$$

Thus, we get:

$$\sigma_i^2(t) = \frac{A_{0i}}{A_i(t)}. \quad (7.11)$$

In this case, the series $\{\sigma_i^2(t)\}_{t \geq t_{0i}}$ and $\{A_i^{-1}(t)\}_{t \geq t_{0i}}$ are proportional and, *a fortiori*, correlated. Therefore, we discard every detected fusion event for which the (empirical) correlation is insufficient, by hypothesis testing. To estimate $\sigma_i^2(t)$ and $A_i(t)$, we resort to a Gauss-Newton algorithm to fit a Gaussian spot in each frame of \widehat{z}_i . Then, to recognize diffusion processes, we rely on the Spearman’s rank correlation coefficient ρ , which is more robust to outliers than Pearson’s test [McDonald, 2014]. The competing hypotheses are:

- H_0 : σ_i^2 and A_i are not correlated;
- H_1 : σ_i^2 and A_i are correlated.

Algorithm 1 Detection and estimation procedure for the point source model

-
- ▷ *Fusion event detection*
 - Select detection scale s^* in I
 - Detect spots e_i in $I_{\delta t}$ at scale s^*
 - for each** e_i **do**
 - Estimate foreground \widehat{z}_i
 - ▷ *Non-diffusing event discarding*
 - for each** $t \in [t_{0i}, t_{0i} + 4]$ **do**
 - Fit a Gaussian spot model to $\widehat{z}_i(t)$
 - if** $\{\sigma_i^2(t)\}_{t \geq t_{0i}}$ and $A_i(t)$ are not correlated **then**
 - Discard e_i
 - ▷ *Fusion model estimation*
 - Estimate model parameters p_{0i}, A_i, σ_i and D_i in $\widehat{z}_i(t_{0i})$
-

The test statistics for the correlation is $t_s = \sqrt{3\rho^2}/\sqrt{1-\rho^2}$, which is t-distributed with three degrees of freedom under H_0 [McDonald, 2014]. Hence, we can infer a threshold on ρ from a given rate of type I error α . As mentioned above, we use 5 consecutive frames to estimate the diffusion coefficient. This results in a threshold on ρ of -0.9 for a rate of type I error of $\alpha = 5\%$ [McDonald, 2014], which we have chosen. We obtained better results with this correlation test than with the variance increase test proposed in [Cortes and Amit, 2008].

The whole proposed detection procedure is summarized in Algorithm 1.

7.2.3 Quantitative evaluation of the diffusive event detection

To evaluate the ability of the method to only retain diffusion events, experiments were conducted on the TIRFM sequence introduced in Figure 7.7. 90 fusion events were detected and then manually labeled as diffusion, non-diffusion, or debatable if the class was unclear.

Table 7.1 – Diffusive event detection performance

α	N	#deb	#diff	#n-diff	#miss	#err
1%	90	6/90	13/34	49/50	21/34	1/50
5%	90	6/90	18/34	48/50	16/34	2/50
10%	90	6/90	20/34	47/50	14/34	3/50

α : rate of type I error; N : detected fusion events; #deb: debatable events; #diff: diffusion events (found/ground-truth); #n-diff: non-diffusion events (found/ground-truth); #miss: missed diffusion events; #err: non-diffusion events erroneously labeled as diffusion.

As summarized in Table 7.1, very satisfactory precision results are obtained (precision ranges from 87% to 92% depending on α). To estimate the diffusion coefficient, it is important to obtain as few erroneous detections as possible, in order to limit the number of irrelevant measurements. Recall is not crucial here. The few classification errors are due to a second spot moving inside v_i , thus affecting the Gaussian fit. As a tradeoff between precision and recall, we will fix the error rate to $\alpha = 5\%$. However, as shown in the next chapter, the choice of α is not critical.

7.2.4 Detection from pH-sensitivity for Rab11 with Dual View

When both Rab11 and a pHluorin-tagged protein are imaged using Dual View, another detection method than the translation/diffusion classification can be employed. Indeed, when detecting an event e_i in the pHluorin channel, we directly obtain the space-time coordinates $e_i = (p_{0i}, t_{0i})$, which are indeed valid in the Rab11-mCherry channel as well.

Also, this method does not rely on any particular dynamical model, so it is compatible with any diffusion model for both imaged proteins, while the classification proposed in Section 7.1 does not handle Rab11 dissociation, for example.

7.3 Discussion

We have proposed two main ways of detecting diffusion dynamics. The first approach consists in classifying vesicle dynamics into translation or diffusion, so the fusion event starts at the first frame where the vesicle dynamics are classified as diffusion. While well suited to Rab11 under the point source model, this solution is not adequate for proteins observed with pH-sensitive tags, since the translation is not visible in that case. Also, as it is specific to the point source model, another approach must be used for the SSED model.

Fortunately, fusion event detection is much easier for TfR and Langerin thanks to the properties of pHluorin, since a fusion event corresponds to a spot appearance in the image sequence. Hence, our spot detector ATLAS is applied to the sequence of temporal differences to locate fusing vesicles.

When working with the point source model instead of the SSED model, a selection is performed among the detected spot appearance events, relying on a statistical test. This allows to discard non-point source events to later obtain reliable estimates of the diffusion coefficient.

Finally, using Dual View, this simple method enables the study of Rab11 diffusion after detecting the fusion events in the TfR or Langerin channel.

8

Estimation of the biological parameters

IN THIS CHAPTER, we propose methods to estimate the biological parameters of the previously introduced translation, point source and SSED models.

First, in order to classify Rab11 dynamics with the method presented in Section 7.1, we propose an estimation method for the translation and point source models which can be performed over a short temporal interval. As explained in the previous chapter, using only three consecutive frames, the method enables to decide the type of dynamics undergone by the vesicles on a frame-by-frame basis.

Then, methods requiring a small temporal integration (5 frames) are presented to estimate the parameters of the point source model. In addition to existing estimation methods, we propose new ones and compare them, in order to select the best method. Quantitative results for the point source model show the superiority of the intensity fitting method, which will then be extended to the SSED model. Since the model is more complex, several improvements are introduced to better fit the data. We will employ this method to compare the dynamical behaviors of TfR and Langerin in real TIRFM image sequences. Specifically, distributions of the residence time τ and diffusion coefficient D will be compared for the two transmembrane proteins.

The chapter organization is the following. Existing methods for estimating the diffusion coefficient are presented in Section 8.1. We propose a method to estimate vesicle drift (translation) in Section 8.2, and to estimate the point source model in Section 8.3. Experimental results show that our method performs best. As for transmembrane dynamics tackled in Section 8.4, the best estimation method for the point source model is improved to cope with the increased complexity. Then, we carry out several experiments which allow us to exhibit differences in the dynamical behaviors of TfR and Langerin. Finally, estimation methods are discussed in Section 8.5, where we also propose improvements for future experiments.

Table 8.1 – Requirements of the diffusion estimation methods

Method	Observed quantity	Source distribution
Particle tracking	Particle trajectory	Sparse
Correlation fitting	Local concentration	Uniform
Intensity fitting	Local concentration	Known

8.1 Existing diffusion estimation methods

Numerous methods were proposed to estimate the diffusion coefficient for all the existing models presented in Chapter 6. Nevertheless, three classes of methods can be distinguished:

- Methods based on single particle tracking (SPT);
- Fluorescence correlation spectroscopy (FCS), which relies on the spatial and/or temporal intensity correlation between spatially and/or temporally neighboring pixels;
- Intensity fitting methods in which an intensity model is formulated and estimated in a space-time volume of the microscopy image sequence.

Hypotheses on the diffusion models required by those methods are summarized in Table 8.1. Let us briefly describe each type of method to enlighten the limitations when applied to the modeling of vesicle fusion.

8.1.1 Single particle tracking methods

Here, we present the basic mean square displacement (MSD) approach to show SPT methods are inappropriate to study vesicle fusion. Nevertheless, more complex SPT models were proposed in the literature and were successfully applied to real images under more realistic hypotheses, although not at vesicle fusion. In particular, authors of [Schuss et al., 2007, Singer et al., 2008] propose to relax the planar membrane assumption, and to address the narrow escape issue introduced in the previous chapter. To our knowledge, however, it was never applied to vesicle fusion, because the followed protein cannot be selectively located in a fusing vesicle. As opposed to SPT methods, we will see that our SSED model relies on higher-level hypotheses to cope with this issue in the context of vesicle fusion, by incorporating a continuous release rate.

SPT-based methods require images where particles can be individually detected. By detecting and tracking moving particles, the diffusion coefficient is generally recovered from the so-called MSD [Sako and Kusumi, 1994, Saxton and Jacobson, 1997, Kim et al., 2011]. MSD is a function of a time step Δt , defined as the expected displacement

of a particle i in an interval step Δt :

$$\text{MSD}(\Delta t) = \left\langle \left(\mathbf{k}_i(t) - \mathbf{k}_i(t + \Delta t) \right)^2 \right\rangle_t, \quad (8.1)$$

where $\langle \cdot \rangle_t$ denotes the temporal averaging.

From the Wiener characterization of Brownian motion (6.1), it is established that the MSD linearly increases with time step:

$$\text{MSD}(\Delta t) = 4D\Delta t, \quad (8.2)$$

where D denotes the diffusion coefficient. Hence, if individual proteins are tracked over time, the diffusion coefficient can be straightforwardly estimated from empirical MSD measurements.

More recently, Hozé *et al.* [Hozé *et al.*, 2012] proposed a method for estimating the diffusion coefficient with a Eulerian approach. After tracking particles, statistical arguments are proposed to characterize the dynamical properties owing to a probability of passing at each position of a grid. This way, local properties are extracted instead of particle motion parameters.

However, in the images we deal with, single proteins cannot be resolved since they are too close from each other compared to the microscope resolution. MSD cannot be computed as a consequence. Other SPT-based methods are ruled out for the same reason.

8.1.2 Correlation fitting methods¹

Temporal image correlation spectroscopy (TICS) can be employed to estimate the diffusion coefficient of the lateral diffusion model even when single particles are not separated. They were successfully applied to images depicting to the stationary model, where a set of Gaussian spots undergo Brownian motion and/or drift [Ohsugi *et al.*, 2006, García-Sáez and Schwille, 2008, Macháň and Hof, 2010, García-Sáez *et al.*, 2010, Kolin and Wiseman, 2007, Di Rienzo *et al.*, 2014].

In TICS, the following correlation function G is introduced [Sergeev, 2004]:

$$G: \quad \mathbb{R}^2 \rightarrow \mathbb{R} \\ (t, \Delta t) \mapsto \frac{\langle I(p, t) I(p, t + \Delta t) \rangle_p}{\langle I(p, t) \rangle_p^2}, \quad (8.3)$$

where $\langle \cdot \rangle_p$ denotes the spatial averaging.

¹Correlation methods were investigated with Anca Caranfil during her Master internship in the Serpico team at Inria Rennes.

In the stationary model, following the Wiener characterization of the Brownian motion (6.1), G does not depend on t , *id est*:

$$G: \mathbb{N} \rightarrow \mathbb{R}$$

$$\Delta t \mapsto \frac{\langle I(p, t) I(p, t + \Delta t) \rangle}{\langle I(p, t) \rangle^2}, \quad (8.4)$$

where $\langle \cdot \rangle$ is now the *spatiotemporal* averaging. However, while valid in classical membrane diffusion experiments [Sergeev, 2004, Kolin and Wiseman, 2007], we will now show that the time independence assumption does *not* hold for the point source model and, *a fortiori*, for our vesicle fusion SSED model.

To our knowledge, the correlation function (8.3) has never been derived for the point source model. To simplify equations, denote $\sigma^2(t) = 2Dt + \sigma_0^2$. The numerator of (8.3) writes:

$$\begin{aligned} \langle I(p, t) I(p, t + \Delta t) \rangle_p &= \frac{1}{|\Omega_I|} \int_{\Omega_I} \frac{A_0}{\sigma^2(t)} \exp\left(-\frac{\|p - p_0\|_2^2}{2\sigma^2(t)}\right) \times \frac{A_0}{\sigma^2(t + \Delta t)} \exp\left(-\frac{\|p - p_0\|_2^2}{2\sigma^2(t + \Delta t)}\right) \mathrm{d}p \\ &= \frac{A_0^2}{|\Omega_I| \sigma^2(t) \sigma^2(t + \Delta t)} \int_{\Omega_I} \exp\left(-\frac{\|p - p_0\|_2^2}{2\sigma^2(t)} - \frac{\|p - p_0\|_2^2}{2\sigma^2(t + \Delta t)}\right) \mathrm{d}p \\ &= \frac{A_0^2}{|\Omega_I| \sigma^2(t) \sigma^2(t + \Delta t)} \int_{\Omega_I} \exp\left(-\frac{\sigma^2(t) + \sigma^2(t + \Delta t)}{2\sigma^2(t) \sigma^2(t + \Delta t)} \|p - p_0\|_2^2\right) \mathrm{d}p \end{aligned}$$

The exponential is negligible when p is not close to p_0 . Thus, the integral over Ω_I is approximately the integral over \mathbb{R}^2 , which is a Gauss integral, and we get:

$$\langle I(p, t) I(p, \Delta t) \rangle_p \approx \frac{\pi A_0^2}{|\Omega_I| (D\Delta t + \sigma^2(t))}.$$

Analogously, the denominator of $G(t, \Delta t)$ is a squared Gauss integral when integrating over \mathbb{R}^2 instead of Ω_I . Then, the following result is straightforward:

$$\langle I(p, t) \rangle_p \approx \frac{2\pi A_0}{|\Omega_I|}.$$

We finally obtain:

$$G(t, \Delta t) \approx \frac{|\Omega_I|}{4\pi (D\Delta t + 2Dt + \sigma_0^2)}, \quad (8.5)$$

which depends on t . In the point source model, non stationarity comes from the pointwise initial particle distribution, which implies that a particle is expected to reach a point $p \neq p_0$ only after a given time which increases with $\|p - p_0\|_2$, thus breaking the stationary assumption for the correlation. The same argument holds for spatiotemporal image correlation spectroscopy (STICS) methods, which include a spatial component in the correlation function (8.3) [Hebert et al., 2005, Kolin and Wiseman, 2007, Di Rienzo et al., 2014].

Therefore, classical FCS methods cannot be applied to estimate the diffusion coefficient for vesicle fusion. However, a new correlation-based estimation method will be proposed in Section 8.3.2 and evaluated in Section 8.3.4.

8.1.3 Intensity fitting method

Intensity fitting was investigated in a few papers to estimate the diffusion coefficient of the stationary [Fortun et al., 2013], point source [Mele et al., 2009] and FRAP models [Seiffert and Oppermann, 2005].

In [Fortun et al., 2013], an optical flow estimation method was proposed to estimate the diffusion coefficient in the stationary model. A global variational approach is used, where a regularization term penalizes high gradients, and a data fidelity term is derived from the Fick's second law. As opposed to correlation fitting which assume a spatially and temporally constant diffusion coefficient, the method proposed in [Fortun et al., 2013] is able to estimate pointwise a varying diffusion coefficient.

Seiffert and Oppermann [Seiffert and Oppermann, 2005] proposed a diffusion coefficient estimation method for the FRAP model, based on frame-by-frame Gaussian fitting. As presented in Section 6.2.5, in the FRAP model, the fluorescence hole has a Gaussian profile of linearly increasing variance $\sigma^2(t)$ and decreasing amplitude $A(t)$:

$$I(p, t) = A(0) - A(t) \exp\left(-\frac{\|p - p_0\|_2^2}{2\sigma^2(t)}\right) \quad (8.6)$$

$$\text{with } A(t) = \frac{A_0}{2D(t - t_0) + \sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2} \quad (8.7)$$

$$\text{and } \sigma^2(t) = 2D(t - t_0) + \sigma_{\text{PSF}}^2 + \sigma_{\text{beam}}^2. \quad (8.8)$$

After fitting a Gaussian spot model in each frame $I(t)$, the amplitude model (8.7) and variance model (8.8) are fitted to recover the diffusion coefficient D , that is merely the slope of A^{-1} and σ^2 . We will see later that best estimation results are obtained with our estimation method.

Finally, Mele *et al.* [Mele et al., 2009] proposed an approach for estimating the diffusion coefficient for vesicle fusion. Intensity is spatially averaged around fusion events to get a 1D+time signal whose decay can be fitted. Specifically, the averaging is performed in a region whose size depends on σ_{PSF}^2 , which must be estimated beforehand. The final estimate of D hence depends on two complex fitting steps. It turns out that much information is lost and the two-step estimator is lacking accuracy.

In the remaining of this chapter, we propose and evaluate a method which can be used in any situations where an intensity model can be derived. In particular, the proposed SSED model will be estimated to better analyze membrane diffusion after vesicle fusion in real TIRFM sequences.

8.2 Proposed estimation method for the translation

Vesicle motion is quite slow – a few pixels per frame at most. Moreover, since we observe only the membrane region, the spatial density of vesicles is low enough to neglect about potential crossings of vesicles. Therefore, instead of using complex methods such as particle tracking (for a survey, see [Chenouard et al., 2014]), it is sufficient to rely on local techniques of optical flow estimation to recover their frame-by-frame displacement. Namely, after detecting the vesicles, we merely estimate the displacement with a least squares method, exploiting the brightness constancy constraint equation (BCCE), which is valid since photobleaching is negligible in the short time interval considered (the frame period). BCCE leads to the well-known motion equation [Lucas and Kanade, 1981]:

$$I_t(p, t) + I_x(p, t)u_i + I_y(p, t)v_i = 0, \quad (8.9)$$

where I_t , I_x and I_y are the temporal and spatial derivatives, and u_i and v_i represent the x and y components of the displacement \mathbf{w}_i of every point of the vesicle, respectively.

The temporal and spatial derivatives of the image intensities must be estimated at each point of the vesicle support. While simple, the finite difference approach is very noise-sensitive. Therefore, we first apply a low-pass filtering with a Gaussian kernel, whose variance is the previously selected scale s^* (see Part I).

The velocity vector $\mathbf{w}_i = (u_i, v_i)^T$ of the vesicle of spatial support $V_i^*(t)$ is easily obtained [Lucas and Kanade, 1981]:

$$\mathbf{w}_i = - \begin{pmatrix} \sum_{q \in V_i^*(t)} I_{xx}(q, t) & \sum_{q \in V_i^*(t)} I_x(q, t) I_y(q, t) \\ \sum_{q \in V_i^*(t)} I_x(q, t) I_y(q, t) & \sum_{q \in V_i^*(t)} I_{yy}(q, t) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{q \in V_i^*(t)} I_x(q, t) I_t(q, t) \\ \sum_{q \in V_i^*(t)} I_y(q, t) I_t(q, t) \end{pmatrix}, \quad (8.10)$$

where I_{xx} and I_{yy} are the second spatial derivatives of the image intensity.

8.3 Proposed estimation method for the point source model

8.3.1 Frame-by-frame estimation by exploiting the Fick's second law

The diffusion coefficient D_i of the vesicle of support $V_i^*(t)$ can straightforwardly be estimated from the Fick's second law (6.12), with a linear least square fitting. In order to estimate D_i more robustly, we do not only consider the points of $V_i^*(t)$ at time t , but also their corresponding points in the previous and next frames, respectively at times $t - 1$ and $t + 1$:

$$D_i = \frac{\sum_{q \in V_i^*(t)} \sum_{\tau=t-1}^{t+1} \Delta I(q, \tau) I_t(q, \tau)}{\sum_{q \in V_i^*(t)} \sum_{\tau=t-1}^{t+1} (\Delta I(q, \tau))^2}. \quad (8.11)$$

We also apply beforehand the aforementioned low-pass Gaussian filter of variance s^* to smooth the Laplacian field, which is very sensitive to noise. This filtering does not impact the estimation, as the Fick's second law is invariant to Gaussian convolution.

8.3.2 Correlation fitting method

We have shown in Section 8.1.2 that classical STICS methods cannot be applied to estimate the diffusion coefficient of vesicle fusion models. However, we propose a simple correlation-based method to estimate the diffusion coefficient under the point source model. From (8.5), we can infer that G^{-1} is a linear function of D and σ_0^2 as:

$$G^{-1}(t, \Delta t) \approx \frac{8\pi}{|\Omega_I|}Dt + \frac{4\pi}{|\Omega_I|}D\Delta t + \frac{4\pi}{|\Omega_I|}\sigma_0^2. \quad (8.12)$$

Interestingly, the result is independent of A_0 and p_0 , so that D and σ_0^2 can be estimated with a standard two-dimensional linear regression. This method will be evaluated in Section 8.4.1.

Unfortunately, we were not able to extend this method to the SSED model, whose correlation function is much more complex. Indeed, non-stationarity is even reinforced in the SSED model by the time-varying source, which introduces another dependence on t . Moreover, those correlation-based methods make the assumption that every particle undergoes a Brownian motion. This hypothesis cannot be assumed in the SSED model because a stationary component is introduced, which prevents FCS from correctly estimating the related parameters. We must resort to intensity fitting methods.

8.3.3 Intensity fitting method

Since we have already estimated $A_i(t)$ and $\sigma_i^2(t)$, as explained in Chapter 7, to detect fusion events, D_i could be straightforwardly deduced from (7.9) or (7.10), as proposed in [Seiffert and Oppermann, 2005] for FRAP experiments. However, as demonstrated in the next section, better results are obtained by directly fitting the diffusion model equation (6.23) to the estimated foreground \widehat{z}_i in the space-time cuboid v_i (see Chapter 7). Specifically, the parameter vector $\theta_i = (p_{0i}, A_{0i}, \sigma_{0i}, D_i)^T$ is estimated by using a Gauss-Newton algorithm initialized with $\theta_i^{(\text{init})} = (\widehat{p}_{0i}, s^* \max_{v_i} \widehat{z}_i, \sqrt{s^*}, 0)^T$:

$$\widehat{\theta}_i = \arg \min_{\theta_i} \sum_{(q,t) \in v_i} r_i^2(q, t) \quad (8.13)$$

$$\text{with } r_i(q, t) = \widehat{z}_i(q, t) - A_i(t) \exp\left(-\frac{\|q - p_{0i}\|_2^2}{2\sigma_i^2(t)}\right).$$

8.3.4 Quantitative comparison of estimation methods²

To carry out an objective and comparative evaluation, we have simulated a collection of 300 image sequences containing a single diffusing spot. Each sequence is made of five images of size 21×21 pixels to mimic real spatiotemporal patches v_i . As explained in Chapter 2, the noise in real TIRFM images is supposed to be Poisson-Gaussian, so the noise variance can be stabilized to end up with a Gaussian noise. We can work with simulated sequences corrupted by Gaussian noise. We create sequences with a random signal-to-noise ratio (SNR) between 1 and 10. The diffusion coefficient and the initial spot variance are randomly varied from 0.1 to 10 px^2/frame , and from 0.5 to 1.5 px^2 , respectively. As shown in Section 8.4.2, diffusion coefficients are of the order of magnitude of 1 px^2/frame in TIRFM image sequences we are dealing with.

We have compared our results with the aforementioned diffusion coefficient estimation methods, namely:

- Direct use of Fick's second law presented in Section 8.3;
- Amplitude decay fitting based on (7.9);
- Variance increase fitting based on (7.10);
- Correlation fitting presented in Section 8.3.2.

Comparative results are reported in Figure 8.1. This clearly demonstrate that our new method outperforms the others. Overall, the estimation of the diffusion coefficient was improved, especially for fast diffusion processes ($D \geq 1 \text{ px}^2/\text{frame}$). This order of magnitude corresponds to real situations, as shown in Section 8.3.4. The mean absolute logarithmic errors (MALEs) of the five methods are respectively:

- 0.24 for the direct least-square estimation,
- 0.22 for the amplitude fitting,
- 0.06 for the variance fitting,
- 0.07 for the correlation fitting,
- 0.03 for our intensity fitting method.

Quantitative evaluation on real TIRFM image sequences

We report now results obtained in real sequences. Six TIRFM image sequences were acquired, depicting micropatterned M10 cells, as in Figure 7.6. TfR was tagged with pHluorin. For each sequence, 600 images of size 256×256 were acquired at 10 f/s. The overall processing time is less than 0.1 s/f on a 2.3 GHz 4-core laptop.

As reported in Figure 8.2a, varying the rate of type I error does not impact significantly the estimated diffusion coefficient statistics. For $\alpha = 5\%$, 124 diffusion events were detected through the six real sequences. The average diffusion coeffi-

²Correlation fitting was evaluated with Anca Caranfil during her Master internship in the Serpico team at Inria Rennes.

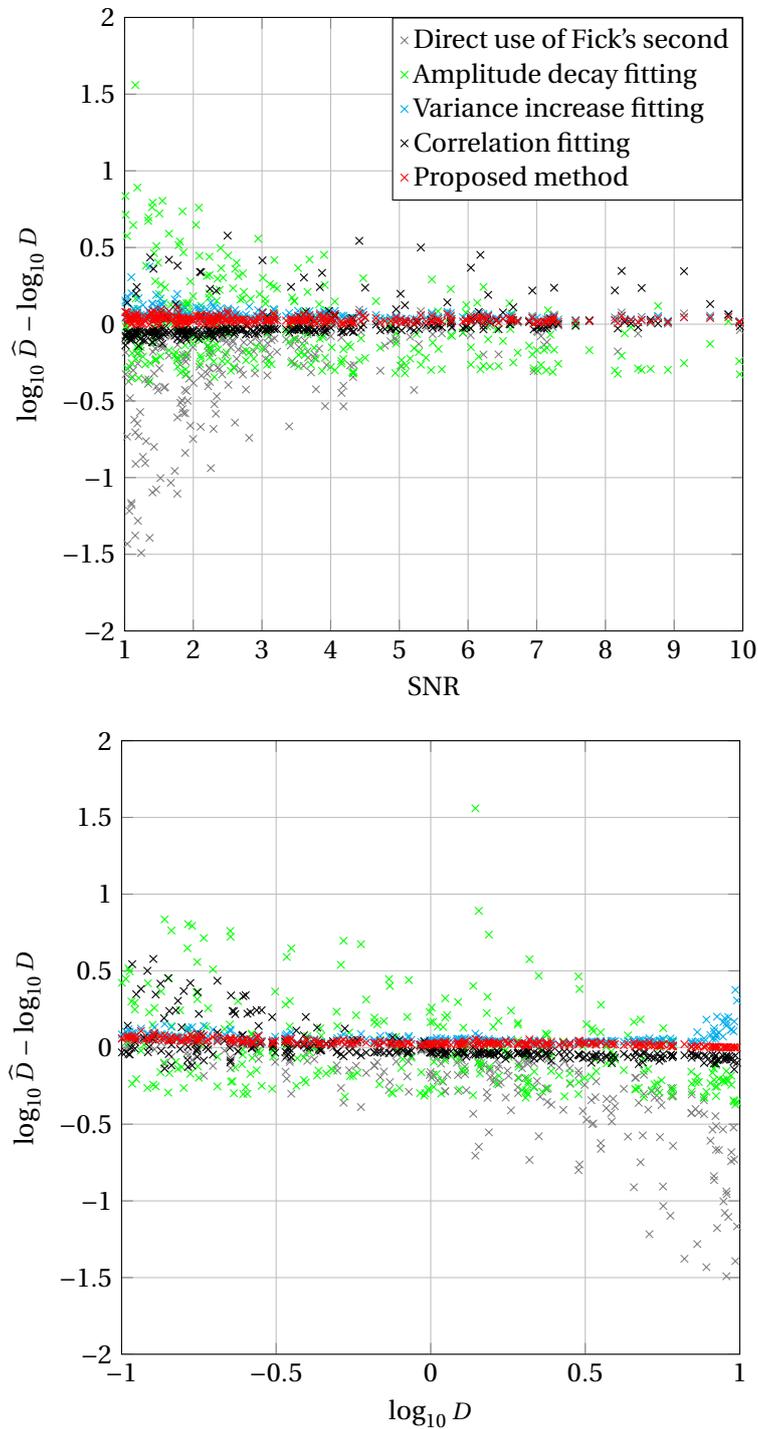


Figure 8.1 – Comparison of the performance of several estimation methods for the point source model. The residuals $\log_{10} \hat{D} - \log_{10} D$ are plotted, where D and \hat{D} are the true and estimated diffusion coefficients, respectively.

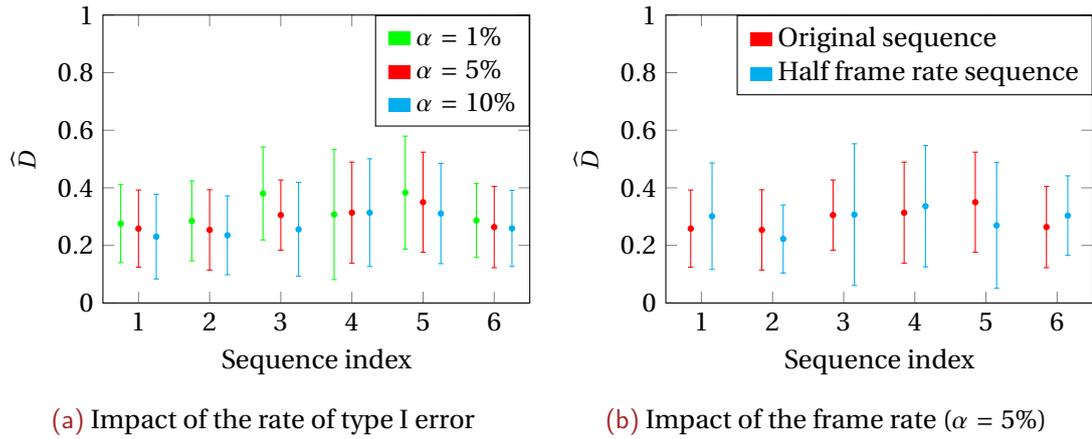


Figure 8.2 – Mean (dot) and standard deviation (bar) of the diffusion coefficient estimated in six real TIRFM sequences

cient equals $1.1 \text{ px}^2/\text{f}$. It converts to $0.28 \text{ }\mu\text{m}^2 \text{ s}^{-1}$, which is coherent with related studies [Ohsugi et al., 2006, Sako and Kusumi, 1994].

Finally, in order to estimate the robustness of the method, we have artificially reduced the frame rate by using only one frame over two. We use the same parameters (detection parameters, size of v_i , and α) as before. Estimation results are presented in Figure 8.2b. Again, the average diffusion coefficient is hardly modified, demonstrating the robustness of the proposed method.

However, the method is only valid for the point source model. Indeed, events with long residence time τ are discarded from the estimation, as explained in the previous chapter. In order to obtain statistics of D including fusion events with large τ , we will propose an extension of this estimation method to the SSED model.

8.4 Proposed estimation method for the SSED model

Let us now focus on the estimation of the SSED model parameters. The model is defined by (6.40), which has no closed form, but can be numerically solved, so that standard optimization algorithms could be used. However, the SSED model has one more parameter than the point source model, and we were not able to satisfyingly estimate the SSED model parameters in simulated sequences using the estimation procedure presented in Section 8.3.3. We need to design a more elaborate algorithm, described below..

First, a good approach is to estimate $\kappa = \tau^{-1}$ instead of τ , as it “reduces non-linearity” in (6.40). Since the Gauss-Newton algorithm does not always converge to the global minimum, we adopt the Levenberg-Marquardt algorithm and the update scheme of [Nielsen, 1999].

Algorithm 2 Detection and estimation procedure for the SSED model

```

for each  $e_i$  do
  Estimate foreground  $z$ 
  Estimate Gaussian spot parameters  $p_{0i}, A_i$  and  $\sigma_i$  in  $\widehat{z}_i(t_{0i})$ 
  Estimate  $\kappa_i$  and  $D_i$  with various  $\theta_i^{(\text{init})}$  in  $\widehat{z}_i$ 
  Retain best fit parameters
  if goodness-of-fit is too low then
    Discard  $e_i$ 

```

Moreover, as the intensity model at $t = t_0$ is a Gaussian spot, we can reliably estimate p_0, A_0 and σ_0 by fitting a Gaussian spot model in frame $I(t_0)$. This way, κ and D can be estimated with a regression operating in two dimensions only.

Finally, the initialization of the parameters, and in particular the initialization of κ , is crucial. Quantitative results show that the estimation of all parameters is accurate when κ is correctly initialized. Therefore, we propose another extension to the estimation procedure. Instead of estimating the parameters only once for each detected event, we start with different initialization vectors, and after running the optimization algorithm, we compare the associated residuals. To obtain the best estimation of κ and D , we select the run which minimizes the sum of squared residuals. In practice, as a tradeoff between accuracy and computation time, we have chosen the set $\{0.1, 0.31, 1, 3.1, 10\}$ of initial values for $\kappa^{(\text{init})}$ and $\{0.1, 10\}$ for $D^{(\text{init})}$.

The overall estimation algorithm is given in Algorithm 2.

8.4.1 Quantitative evaluation on simulated sequences

The same procedure as for the point source model evaluation is used to evaluate the proposed estimation method for the SSED model. 300 synthetic image sequences of size 21×21 pixels and length 20 frames were generated with different parameters to mimic real \widehat{z}_i 's. Like for the point source model, we have randomly set the diffusion coefficient between 0.1 and $10 \text{ px}^2/f$, and the PSF variance from 0.5 to 1.5 px^2 . As for the residence time τ , it varies between 0.1 and 10 frames. The SNR still ranges from 1 to 10.

Logarithmic errors on the estimation of both κ and D are reported in Figure 8.3 for each sequence. As plotted in Figure 8.3a, the estimation of κ is less accurate than that of D , but we will see in the next subsection that the accuracy is largely sufficient to extract relevant information from real TIRFM images. Moreover, large errors are very rare. Over the 300 generated sequences, only 5 have an absolute logarithmic error higher than 0.5, and the MALE is quite low at 0.12.

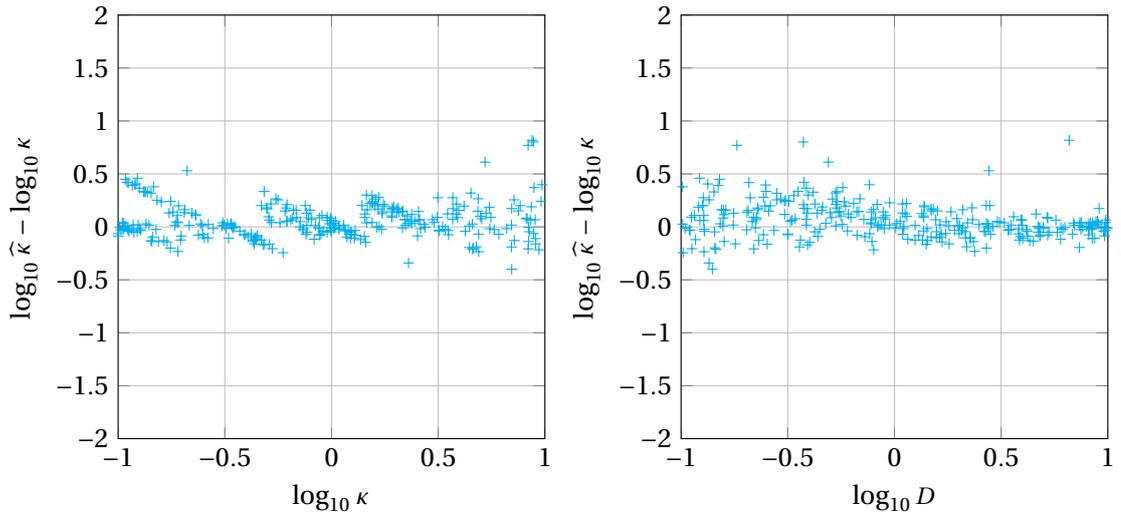
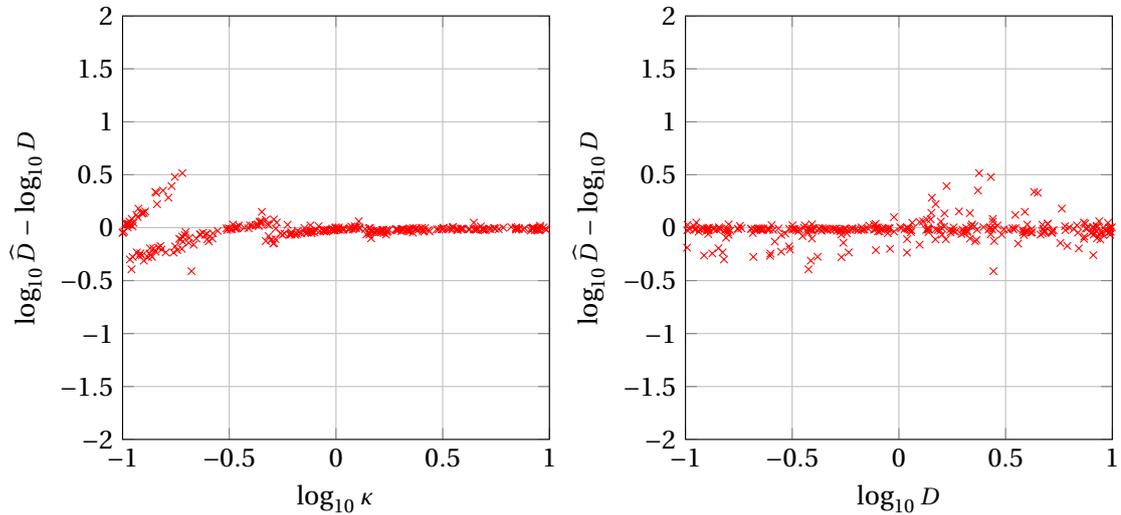
(a) Accuracy of the estimation of κ versus the true parameter values(b) Accuracy of the estimation of D versus the true parameter values

Figure 8.3 – Results obtained for the SSED model on simulated sequences

As for the estimation of D , results reported in Figure 8.3b are very good when κ is high enough. Indeed, this behavior is not a surprise, since, for low κ , the flow between C_s and C_d is very small. Consequently, few particles are available to estimate D (precisely the one undergoing a Brownian motion). On the contrary, when increasing κ , the SSED model tends to the point source model, and the estimation becomes more and more accurate as the amount of signal available to estimate D increases. When $\kappa > 0.25$, estimation of D is as precise as for the point source model with a MALE of 0.03. Including the worst estimates, the overall MALE for the diffusion coefficient is still very low at 0.06.

8.4.2 Comparison of TfR and Langerin behaviors

The behaviors of TfR and Langerin after fusion have been compared in [Cinquin, 2011], but no biophysical model was proposed. Instead, a simple 1D+time intensity signal was used to classify fusion events as *slow* or *fast* in collections of sequences depicting either TfR or Langerin. It was reported that the number of slow events in Langerin image sequences is much higher than in TfR images [Cinquin, 2011].

Similar results can be expected when estimating the SSED model with our method in similar image sequences. In the SSED model, slow events correspond to long residence times (i.e., small κ), while fast events correspond to short residence times (high κ).

We have applied the proposed detection and estimation procedure to sixteen real TIRFM image sequences, half of which depict TfR, and half of which depict Langerin. As explained in Chapter 7, in order to only take into account reliable estimates, a chi-square goodness-of-fit test is performed [McDonald, 2014]. Like for the point source model in previous chapter, we fix the rate of type I error α to 5%. In the set of sequences depicting TfR, 3,147 diffusive events are detected, versus 4,223 for Langerin.

The results are gathered in Figure 8.4 in the form of four histograms of $\hat{\kappa}$ and \hat{D} estimated in the sequences depicting TfR or Langerin. There are two main conclusions to be drawn. First, the histogram of $\hat{\kappa}$ exhibit a strong peak around 10^2 for TfR which does not exist for Langerin. In contrast, much more *slow* events are found in Langerin sequences, around $\hat{\kappa} \approx 10^{-1}$. Regarding the release rate, our results are therefore consistent with those reported in [Cinquin, 2011], but supply a more precise description of the biophysical model with a quantified biophysical parameter.

The second conclusion concerns the diffusion coefficient statistics. Indeed, Langerin shows a much higher dispersion of the estimates than TfR. Various hypotheses could explain such different behaviors, related, for example, to structures present around the vesicle fusion location, interactions between proteins...

These preliminary results demonstrate the relevance of our approach. Statements on κ and D raise biological questions, answers of which could lead to a better understanding of the structures and interactions involved in the recycling process.

8.5 Discussion and perspectives

We have designed methods to estimate three types of membrane dynamics using TIRFM, namely translation, 2D and 3D diffusions. First, we developed methods using only three frames to estimate the translation of the vesicle and its diffusion dynamics. While simple and fast, the diffusion coefficient estimation is not sufficiently accurate to draw reliable conclusions on real TIRFM image sequences.

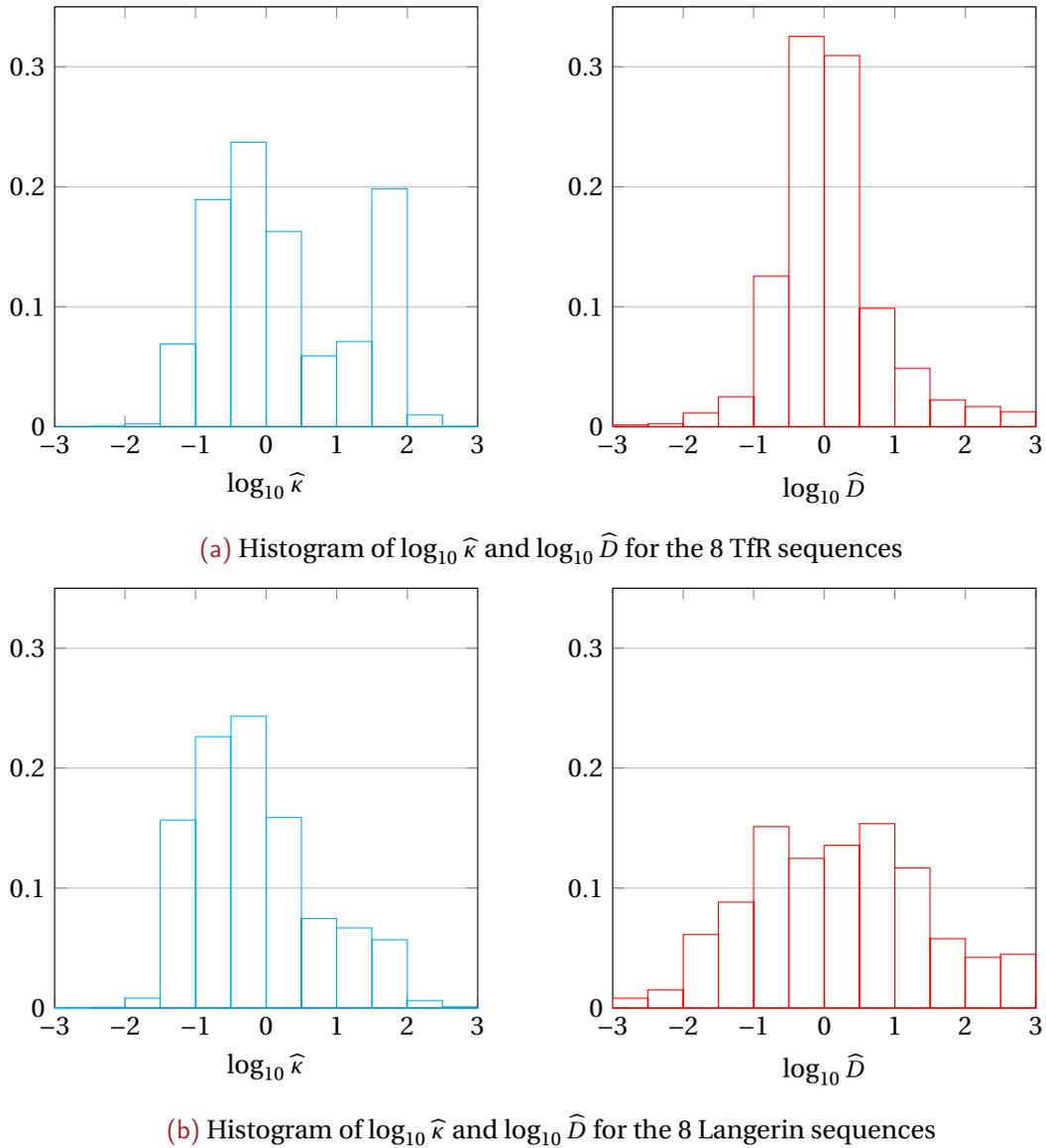


Figure 8.4 – Comparison of the histograms of the biophysical parameters $\hat{\kappa}$ and \hat{D} estimated in 16 TIRFM image sequences depicting TfR or Langerin

As a consequence, we investigated a more efficient approach for the point source and SSED models, relying on a longer temporal integration (5 and 20 frames, respectively). After demonstrating the efficiency of the method on simulated sequences, we successfully applied it to real TIRFM images depicting TfR and Langerin.

The experiments demonstrated that the residence time and diffusion coefficient distributions of the two transmembrane proteins exhibit clear differences. A biological model to explain it remains to be elaborated. In collaboration with UMR 144, further work could focus on the design and validation of biological models which could explain the behavior differences. The estimation of the SSED model should also be carried out in sequences showing mutant and wild-type cells.

III

Group dynamics

Labeled affine flow for group motion characterization

GROUP DYNAMICS can be encountered in biological image sequences, for example images of cell tissues as depicted in Figure 9.1, embryogenesis, cell migration, and in many other types of image sequences, ranging from urban traffic surveillance to crowd motion monitoring. The latter is particularly interesting in terms of evaluation because many crowd motion analysis methods have been proposed in the literature. Therefore, this scope represents an excellent experimental support.

Analyzing videos of crowded scenes is of interest in many applications [Zhan et al., 2008, Chandola et al., 2009, Thida et al., 2013]. Needs can encompass crowd safety in big social events, exhibitions, sports events or musical shows, surveillance and monitoring in public transportation areas like subways, airports or railway stations, people behavior understanding in commercial venues. The huge amount of daily acquired videos urges to define automatic tools for processing the available data and/or assisting human operators depending upon the targeted applications. Goals can comprise classifying dynamic behaviors [Hu et al., 2008, Rodriguez et al., 2011, Solmaz et al., 2012, Zhou et al., 2013], identifying main followed paths [Jodoin et al., 2013, Wang et al., 2011, Zhou et al., 2011], preventing critical situations [Solmaz et al., 2012], and detecting abnormal behaviors or events [Cong et al., 2013, Kim and Grauman, 2009, Kratz and Nishino, 2009, Wu et al., 2014].

A similar image analysis process can be envisaged as well for other sets of moving elements such as urban traffic, animal flocks or groups of cells like those depicted in Figure 9.1. However, in contrast to most existing methods, our methods are said instantaneous, that is, they rely on short-term analysis and do not require any prior learning stage. More precisely, the proposed methods are based on simple affine motion models estimated from only two consecutive images. They do not involve long temporal integration or trajectory computation. Furthermore, in order to create methods as generic and reusable as possible, we do not introduce any individual ap-

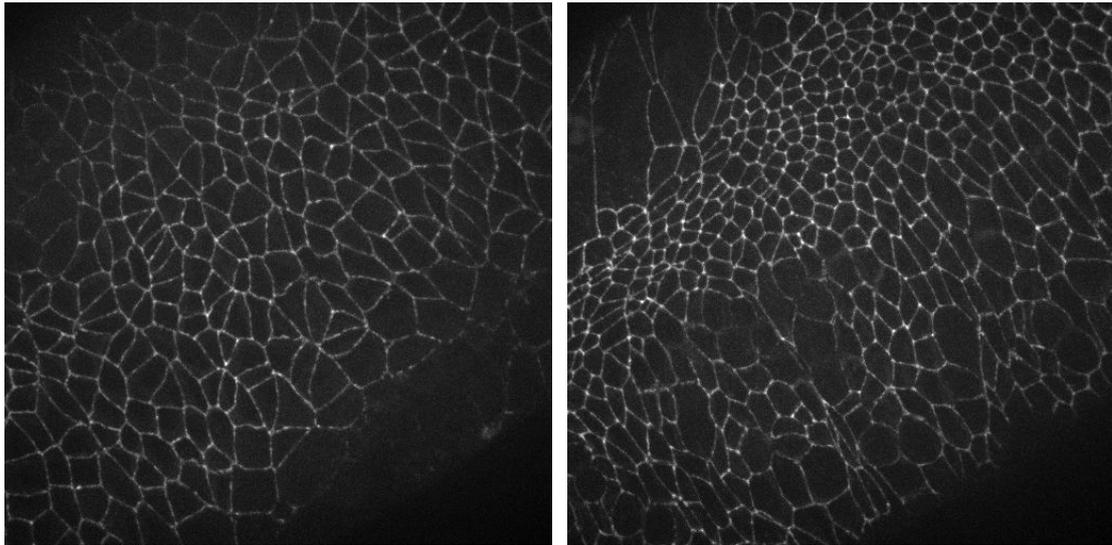
(a) First frame $I(0)$ of the sequence(b) Last frame $I(199)$ of the sequence

Figure 9.1 – Tissue sequence depicting cells whose plasma membranes are fluorescently labeled

pearance or interaction model. While applicable to biological images showing groups of cells, the proposed methods do not rely on hypotheses specific to biology, so that they can be applied to any kind of microscopy images, and many other application domains.

In this chapter, we are introducing labeled affine flow (LAF), a map incorporating both a quantitative motion information, in the form of displacement vectors, and a qualitative motion label, in the form of motion classes. The term affine flow indicates that the motion is estimated with affine motion models. Indeed, we assume that the apparent motion of a group can locally be represented by one of the three following 2D motion models: translation, scaling or rotation, which are three specific cases of the affine motion. Scaling motions correspond to gathering (*Convergence*) or dispersing (*Divergence*) when viewed by the camera. Rotation motions are subdivided into *Clockwise* and *Counterclockwise* classes. Since our classification scheme is view-based, we choose to distinguish four image-related translation directions: *North*, *West*, *South*, *East*. A finer subdivision could be handled as well if required.

These eight group motion classes can be related to the *behaviors* introduced in [Solmaz et al., 2012], as summarized in Table 9.1. However, let us mention that our scheme is applicable to any point in the image, not only around few critical points as in [Solmaz et al., 2012].

Our motion estimation method is divided into three main steps:

1. Detection of moving areas in the image;
2. Estimation of three affine motion models (translation, rotation, scaling) in each window of a collection of windows of different sizes;

Table 9.1 – Group motion classes compared to behaviors from [Solmaz et al., 2012]

Motion type	Motion class direction	Behavior from [Solmaz et al., 2012]
T	North West South East	Lane
S	Convergence Divergence	Bottleneck Fountainhead
R	Clockwise Counterclockwise	Ring

3. Pointwise selection of the optimal motion models.

The first step is achieved with the motion detection algorithm [Crivelli et al., 2011], which follows a background subtraction approach and involves a mixed-state conditional random field (see [Crivelli et al., 2011], for details). The set of detected moving areas is denoted Ω_I^* with $\Omega_I^* \subset \Omega_I$, where Ω_I is the image domain. The motion detection algorithm is tuned by two parameters that we have kept fixed for all processed sequences, and their setting was not critical for our classification task. With step 2, we end up with a set of motion model candidates at every point $p \in \Omega_I^*$. Step 3 allows us to select at every point $p \in \Omega_I^*$ the most relevant motion model among these candidates with an information criterion.

The outline of the chapter is the following. After giving a brief overview of crowd motion analysis methods in Section 9.1, the proposed LAF map is introduced in Section 9.2. Two applications are then proposed to demonstrate the relevance of the LAF motion map: motion patterns recognition in Section 9.3 and recovery of dominant paths in Section 9.4. We finally discuss the methods in Section 9.5.

9.1 Related work

Important research efforts have been devoted to crowd analysis for several years [Thida et al., 2013, Li et al., 2015]. Specialized descriptors have been designed to capture the dynamics of crowds motion from videos and have been used for a number of inference tasks in crowd analysis, such as pedestrian tracking [Hu et al., 2008, Rodriguez et al., 2009], group motion pattern or path classification [Wang et al., 2011, Zhou et al., 2011, Solmaz et al., 2012], or anomaly detection [Basharat et al., 2008, Kratz and Nishino, 2009, Feng et al., 2010, Ryan et al., 2011, Chockalingam et al., 2013].

Existing crowd analysis methods usually exploit motion-based features computed on extended time intervals: spatio-temporal cuboids [Feng et al., 2010, Kratz and

Nishino, 2009, Rodriguez et al., 2011], tracklets [Zhou et al., 2011] and mostly trajectories [Cheriyadat and Radke, 2008, Rodriguez et al., 2009, Solmaz et al., 2012, Wang et al., 2011, Zhou et al., 2012]. Moreover, most crowd analysis methods rely on dense optical flows [Chockalingam et al., 2013, Cong et al., 2013, Kim and Grauman, 2009, Mehran et al., 2009, Ryan et al., 2011, Solmaz et al., 2012, Zhou et al., 2013], temporal or spatiotemporal gradients [Benezeth et al., 2011, Boiman and Irani, 2007, Kratz and Nishino, 2009, Roshtkhari and Levine, 2013].

The recognition of prominent paths in the scene is investigated in [Wang et al., 2011] within the framework of hierarchical Dirichlet processes and latent topics, while in [Zhou et al., 2011] random field topic models are introduced which account for spatial and temporal coherence between tracklets. The authors of [Zhou et al., 2013] have proposed a collectiveness measure based on trajectories and local velocities.

Tracking pedestrians in dense crowds may require specific approaches. Idrees *et al.* [Idrees et al., 2014] made the assumption that neighboring pedestrians have comparable trajectories to track individuals in dense crowds. In [Rodriguez et al., 2011], a dense optical flow is first computed to deduce words, composed of quantized positions and directions. Correlated topic models (CTM) are exploited to learn word priors on a large video database. These models are then exploited for tracking individuals in a dense crowd.

As for crowd behavior classification, [Zhou et al., 2012] and [Cheriyadat and Radke, 2008] studied coherent and dominant crowd motions. Zhou *et al.* [Zhou et al., 2012] proposed to group moving points according to the so-called coherent neighbor invariance. The latter provides information both on the spatial proximity of data points and on the correlation over time of their velocity vectors. In [Cheriyadat and Radke, 2008], the trajectories are organized into clusters according to a longest common subsequence (LCSS) criterion. In [Wang et al., 2011], a method is developed for the recognition of semantic regions (prominent paths in the scene) within the framework of hierarchical Dirichlet processes and latent topics, while [Zhou et al., 2011] introduced the so-called random field topic model for semantic region analysis to account for spatial and temporal coherence between tracklets.

To our knowledge, only [Hu et al., 2008] and [Solmaz et al., 2012] have focused on classifying structured group motions. The former determined motion patterns by clustering 4D flow vectors (2D position and velocity of points) in each frame according to proximity and similarity rules. The latter proposed to extract trajectories and accumulation points from the advection of flow fields over video sequences.

9.2 Labeled affine flow

9.2.1 Motion model candidates

As aforementioned, to propose motion candidates, we only consider 2D parametric motion models. Specifically, at any point $p = (x, y) \in \Omega_I^*$, the optical flow vector $\mathbf{w}(p)$ is approximated by an affine flow vector $\mathbf{w}_\theta(p)$ defined by:

$$\mathbf{w}_\theta(p) = \underbrace{\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}}_A \begin{pmatrix} x \\ y \end{pmatrix} + \underbrace{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}}_B, \quad (9.1)$$

with $\theta = (a_1, a_2, a_3, a_4, b_1, b_2)$ the model parameter vector.

In order to characterize the eight previously introduced group motion classes, only three specific affine motion models are necessary: *translation* (denoted T), *scaling* (S), and *rotation* (R) motions. They respectively correspond to the following 2×2 matrices A , as explained in [François and Bouthemy, 1990]:

$$A_T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad A_S = \begin{pmatrix} a_1 & 0 \\ 0 & a_1 \end{pmatrix} \quad A_R = \begin{pmatrix} 0 & a_2 \\ -a_2 & 0 \end{pmatrix}. \quad (9.2)$$

The vector B is considered in any case, since it corresponds to the displacement of the origin of the coordinate system. Hence, for each motion model, only two (translation case) or three coefficients (scaling and rotation cases) have to be estimated, respectively:

$$\theta_T = (b_1, b_2) \quad \theta_S = (b_1, b_2, a_1) \quad \theta_R = (b_1, b_2, a_2). \quad (9.3)$$

Since we do not know in advance the appropriate spatial support to estimate the motion models, we consider a collection $\mathcal{W} = \{W_i | 1 \leq i \leq |\mathcal{W}|\}$ of overlapping windows of various sizes – typically, 25%, 50%, and 100% of the image dimensions. For a given size, the overlap rate is 50%, so that a given point p belongs to four windows of that size (apart from border effects). An example is given in Figure 9.2.

We estimate the three motion models defined in (9.2) in every window, using the robust method [Odobez and Bouthemy, 1995] based on a multiresolution and incremental scheme, which is available as an open-source software¹. The robust estimation allows us to capture the dominant motion if several motions are present inside the window, and to tolerate errors of the motion detection stage. Since the minimization of the robust penalty function amounts to an iteratively reweighted least squares (IRLS) procedure [Odobez and Bouthemy, 1995], each point p is assigned at the end of the estimation process a weight representing its influence in the robust estimation. A

¹Motion2D: <http://www.irisa.fr/vista/Motion2D/>

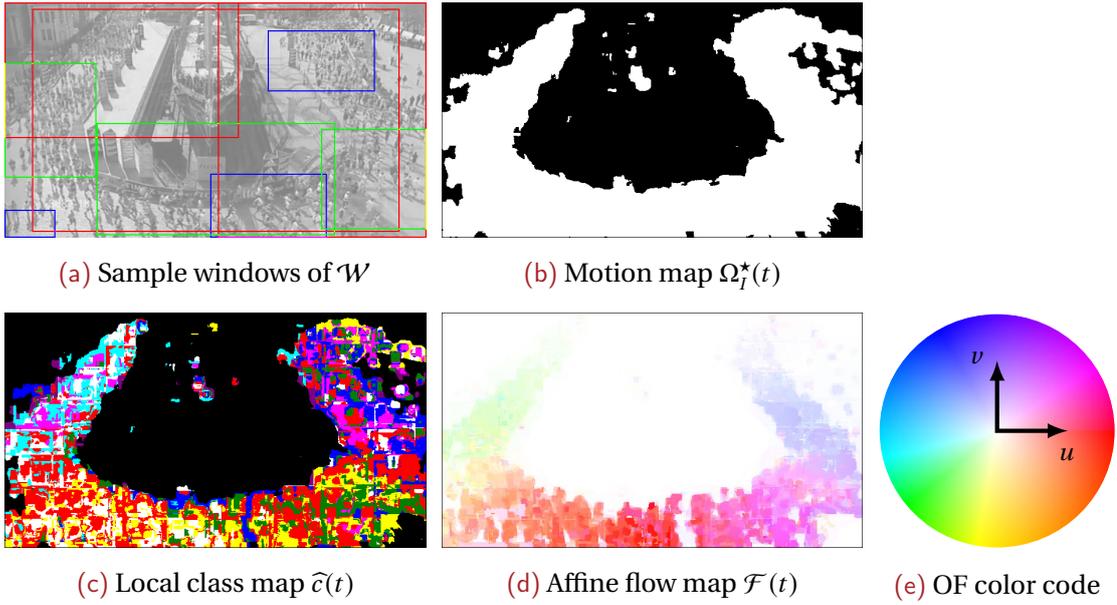


Figure 9.2 – Overview of the motion estimation method applied to the *Marathon bend* sequence, where runners follow a ‘U’ from the upper left corner to the upper right corner. (a) Sample windows from \mathcal{W} are plotted over the first frame of the sequence (red: large windows; green: medium ones; blue: small ones). (b) Map of the moving points detected in the first frame. (c) Motion models are estimated in each window, and candidates are pointwise selected (class color code in Table 9.2). (d) Optical flow deduced from the affine parameters of selected candidates. (e) Color code of (d).

point whose weight is close to 1 (namely greater than 0.5) is called an inlier. Let $\theta_{i,k}$ be the parameters of the motion model $k \in \{T, S, R\}$, estimated in the window $W_i \in \mathcal{W}$. The set of inliers for the model of parameter vector $\theta_{i,k}$ is denoted by $\mathcal{X}_{i,k}$.

The conformity evaluation of a point p to a given motion model of parameters $\theta_{i,k}$ is based on the displaced frame difference (DFD) and is defined by:

$$\varepsilon_{i,k}(p, t) = I(p + \mathbf{w}_{i,k}(p, t), t + 1) - I(p, t), \quad (9.4)$$

where $\mathbf{w}_{i,k}(p, t)$ is the displacement of p between frames t and $t + 1$ deduced from $\theta_{i,k}$ according to (9.1). Conformity corresponds to $\varepsilon_{i,k}$ close to 0.

For every motion model k , in every window W_i , we estimate both the motion parameters $\theta_{i,k}$ and the empirical variance $\sigma_{i,k}^2$ computed over the inliers, given by:

$$\sigma_{i,k}^2 = \frac{1}{|\mathcal{X}_{i,k}|} \sum_{q \in \mathcal{X}_{i,k}} \varepsilon_{i,k}^2(q), \quad (9.5)$$

where $|\mathcal{X}_{i,k}|$ denotes the number of inliers.

Let $\mathcal{W}(p) \subset \mathcal{W}$ be the subset of windows containing a given point p , $\mathcal{M}(p)$ the set of motion model candidates for p , and $\Theta(p)$ the set of estimated parameters of the candidates. In our experiments, using the previously mentioned windows collection,

33 motion model candidates are available for each pixel (only 30 for pixels lying on the image borders).

9.2.2 Candidate selection

The optimal motion model at p should best fit the real (unknown) local motion at p while being of the lowest possible complexity. We consider a local patch $\eta(p)$ centered in p and we exploit the fitting variable (9.4) which is likely to be close to 0 for the correct velocity vector. Let us assume that the $\varepsilon_{i,k}(q, t)$'s are i.i.d. variables over points $q \in \eta(p)$ and follow a zero-mean Gaussian law of variance $\sigma_{i,k}^2$. Then, we can write the joint likelihood in the patch $\eta(p)$:

$$\psi(p, \theta_{i,k}) = \frac{1}{\sqrt{2\pi\sigma_{i,k}^2}^{|\eta(p) \cap \Omega_I^*|}} \prod_{q \in \eta(p) \cap \Omega_I^*} \exp\left(-\frac{\varepsilon_{i,k}^2(q)}{2\sigma_{i,k}^2}\right). \quad (9.6)$$

To penalize the complexity of the motion model, we resort to the Akaike information criterion corrected for small sample size (AICc) [Burnham and Anderson, 2002]. The correction is especially useful when the sample size is small, which is our case here. The criterion is given by:

$$\text{AICc}(p, \theta_{i,k}) = -2 \ln(\psi(p, \theta_{i,k})) + 2n_k + \frac{2n_k(n_k + 1)}{|\eta(p) \cap \Omega_I^*| - n_k - 1}, \quad (9.7)$$

where n_k is the dimension of the motion model k , that is $n_k = 2$ for T-motion model and $n_k = 3$ for S- and R-motion models. Finally, we select the optimal motion model \hat{k} at p by minimizing the criterion:

$$\theta_{i,\hat{k}} = \arg \min_{\theta_{i,k} \in \Theta(p)} \text{AICc}(p, \theta_{i,k}). \quad (9.8)$$

9.2.3 From motion candidates to motion classes

At every point $p \in \Omega_I^*$, we have selected the motion model candidate, which is represented by:

- a motion type $\hat{k} \in \{\text{T, S, R}\}$,
- a parameter vector $\theta_{i,\hat{k}}$ and, consequently, an affine flow vector $w_{i,\hat{k}}(p)$.

The selected candidates are further classified into eight motion classes, represented in Table 9.2 by colors, depending on the sign of the parameters and some combinations of them. Namely, the translation motion type is subdivided into four classes defined from view-based directions:

- Translation towards the top of the frame, or Northward translation (denoted TN);
- Westward translation (TW);

Table 9.2 – Motion classes definition

Motion type	Motion class	Class color	Class direction	Criterion
T	TN	●	North	$b_1 + b_2 > 0 \ \& \ b_1 - b_2 < 0$
	TW	●	West	$b_1 + b_2 < 0 \ \& \ b_1 - b_2 < 0$
	TS	●	South	$b_1 + b_2 < 0 \ \& \ b_1 - b_2 > 0$
	TE	●	East	$b_1 + b_2 > 0 \ \& \ b_1 - b_2 > 0$
S	SN	●	Convergence	$a_1 < 0$
	SP	○	Divergence	$a_1 > 0$
R	RN	●	Clockwise	$a_2 < 0$
	RP	●	Counterclockwise	$a_2 > 0$

- Southward translation (TS);
- Eastward translation (TE).

The scaling motion type corresponds to two classes:

- Convergence, or negative scaling (SN);
- Divergence (SP).

And the rotation motion type naturally yields two classes:

- Clockwise rotation, or negative rotation (RN);
- Counterclockwise rotation (RP).

The different classes are characterized by criteria summarized in Table 9.2. The set of classes is denoted $\Gamma = \{TN, TW, TS, TE, SN, SP, RN, RP\}$, and we finally obtain the local classification map \hat{c} , as depicted in Figure 9.2c.

In the meantime, we also obtain an estimate $\mathbf{w}_{i,\hat{k}}(p)$ of the velocity vector at every point p . The flow map will be denoted by $\mathcal{F}(t)$. We term LAF the combination of

Algorithm 3 Label affine flow estimation

▷ *Motion detection*

Determine the moving regions Ω_i^*

▷ *Motion estimation*

for each $W_i \in \mathcal{W}$ **do**

Estimate the motion parameters $\theta_{i,k}$, $k \in \{T, S, R\}$

▷ *Model selection*

for each $p \in \Omega_i^*$ **do**

for each $\theta_{i,k}$ **do**

Evaluate the variable $\varepsilon_{i,k}(p)$.

Evaluate the likelihood $\psi(p, \theta_{i,k})$.

Select the best model according to the AICc.

the affine flow and motion class maps. The steps for obtaining it are summed up in Algorithm 3.

9.3 Motion patterns recognition

The initial classification map \widehat{c} is noisy, due to the pixelwise selection process, as illustrated in Figure 9.2c. In order to characterize group motions, we propose to regularize \widehat{c} . As we will see in the subsequent subsections, by regularizing the map with simple weighted votes, we also directly obtain the motion patterns at the desired scale. While the voting procedure is lighter than classical regularization approaches, based for example on Markov fields or morphological operations, very good results are reported below. Moreover, the regularization algorithm complexity is linear with the image size and does not depend on the regularization scale.

9.3.1 Local class histograms

For each point $p \in \Omega_I^*$, we first compute a local class histogram $\mathcal{H}(p) = (\mathcal{H}_c(p))_{c \in \Gamma}$ counting the number of points of each class in a neighborhood of p . To avoid block artifacts induced by square neighborhoods, and to put more weight on the pixels q closer to p , we rely on a window function g_r of radius r . Hence, the bin \mathcal{H}_c of class $c \in \Gamma$ is evaluated as:

$$\mathcal{H}_c(p) = \sum_{q \in \Omega_I^*} g_r(\|q - p\|_2) \delta_c(\widehat{c}(q)), \quad (9.9)$$

$$\text{where } \delta_c(\widehat{c}(q)) = \begin{cases} 1 & \text{if } \widehat{c}(q) = c \\ 0 & \text{otherwise.} \end{cases}$$

Typically, we use a Gaussian window of standard deviation r for g_r , and the computation time can be reduced by relying on recursive approximations of the Gaussian filter, such as the Deriche separable filter [Deriche, 1990]. Using this approximation, the computation of the histograms is proportional to the image size, or more precisely the area of Ω_I^* . It is about three times longer than using an unweighted square window.

9.3.2 Class map regularization with inhibition

We propose a two-step method to regularize the local class map \widehat{c} . We first build a regularized motion type map \mathcal{K} from the local type map $estimk$, before inferring the regularized class map \mathcal{C} . This is accomplished with a two-round voting procedure. The notation of the types and classes, and of the associated maps, is summarized in Table 9.3.

Table 9.3 – Classification notation

	Local map	Regularized map	Possible values							
Motion type	\widehat{k}	\mathcal{K}	T				S		R	
Motion class	\widehat{c}	\mathcal{C}	TN	TW	TS	TE	SN	SP	RN	RP

To obtain \mathcal{K} , accumulating votes of the same motion type regardless of the underlying motion class is unavailing and even counterproductive in some situations. As a consequence, points in $\mathcal{V}(p)$ locally assigned with classes *clockwise rotation* and *counterclockwise rotation*, are not considered as bearing corroborating information on rotation but rather contradictory information. The same holds for class pairs *divergence* and *convergence*. Therefore, we introduce the notion of inhibition and reinforcement in the regularization procedure.

The algorithm proceeds as follows. In a given $\mathcal{V}(p)$ at time instant t , we compute the occurrences of the eight group motion classes of Γ from the values $\widehat{c}(q, t)$, $q \in \mathcal{V}(p)$. From the computed occurrences, we build the signed histogram $\mathcal{H}(p, t)$ where bins corresponding to opposite classes (the pairs listed above) have opposite signs, that is, we implement the inhibition factor.

As for translations, the inhibition could be added to the pair of *North translation* and *South translation* classes, and *West translation* and *East translation* classes, respectively. However, in contrast to scaling- and rotation-motion type classes, we rarely observe neighboring opposite directions of the translation-type classes. Adding the inhibition factor to them would be possible but more complex than for scaling and rotation types. Moreover, it would not be very useful as this case is very rare in practice. Conversely, classes of the T-motion type benefit from the reinforcement factor. Nevertheless, non-coherent translations will be tackled in the next chapter.

The first round of the voting procedure consists in selecting the motion type $\mathcal{K}(p)$, as illustrated in Figure 9.3. Three motion type scores \mathcal{H}_T , \mathcal{H}_S and \mathcal{H}_R are computed as follows:

$$\mathcal{H}_T(p) = \mathcal{H}_{TN}(p) + \mathcal{H}_{TW}(p) + \mathcal{H}_{TS}(p) + \mathcal{H}_{TE}(p), \quad (9.10)$$

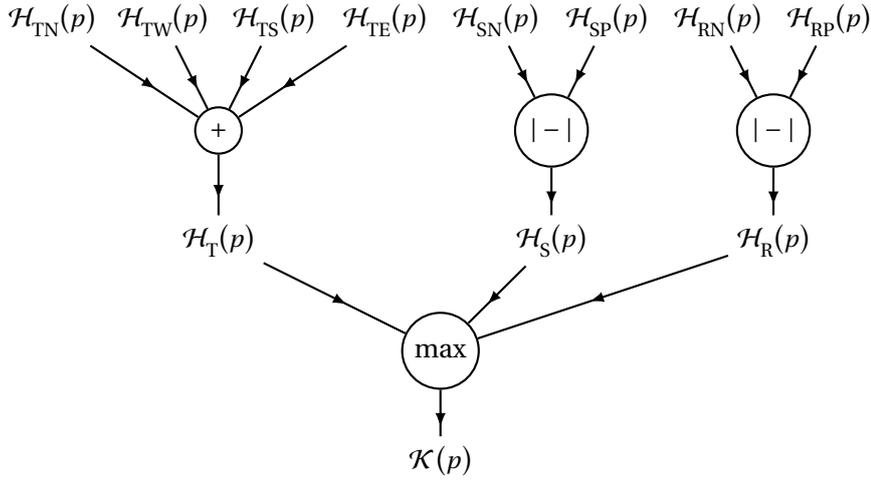
$$\mathcal{H}_S(p) = |\mathcal{H}_{SN}(p) - \mathcal{H}_{SP}(p)|, \quad (9.11)$$

$$\text{and } \mathcal{H}_R(p) = |\mathcal{H}_{RN}(p) - \mathcal{H}_{RP}(p)|. \quad (9.12)$$

The regularized motion type $\mathcal{K}(p) \in \{T, S, R\}$ of maximum score is finally assigned to p :

$$\mathcal{K}(p) = \arg \max_{k \in \{T, S, R\}} \mathcal{H}_k(p). \quad (9.13)$$

The second round of the regularization procedure consists in assigning the group motion class $\mathcal{C}(p)$ which is the most represented among the classes associated to the



$| - |$ represents the absolute difference operator, i.e., the L_1 distance

Figure 9.3 – Classification regularization with reinforcement and inhibition factors

selected motion type $\mathcal{K}(p)$:

$$C(p) = \underset{\substack{c \in \Gamma \\ \text{s.t. } k(c) = \mathcal{K}(p)}}{\arg \max} \mathcal{H}_c(p). \quad (9.14)$$

In turn, we are able to recover the motion patterns of the group behavior on a frame-by-frame basis without any prior learning stage and without any critical parameter setting. Indeed, for an image sequence of τ images, we end up with $\tau - 1$ instantaneous maps $C(t)$ of group motion classes.

9.3.3 Experimental results

We have carried out experiments on real image sequences to validate our new group motion pattern recognition method. Figure 9.4 demonstrates the advantage of penalizing the complexity of the motion models and inhibiting local opposite motion classes. In that example, people do not run perfectly straight ahead from the left to the right of the image, resulting in local irrelevant rotation or scaling decisions.

Classification errors are avoided by using both the AICc (instead of the maximum likelihood (ML) criterion used in our preliminary method [Basset et al., 2013]) and the inhibition procedure. Irrelevant local motions generally comprise opposite motions in that case. As an example, if a runner goes away from someone, he/she comes closer to another neighbor. The tendency would be to classify those neighboring points of \hat{c} as *Convergence* and *Divergence*. In such a case, the inhibition factor helps reducing the score of the scaling motion type, and recovering the correct translation type.

In Figure 9.5, we investigate the impact of the size of the regularization window. Its radius r is varied from 40 to 80 pixels for the *Marathon bend* sequence, where different types of motion can be distinguished.

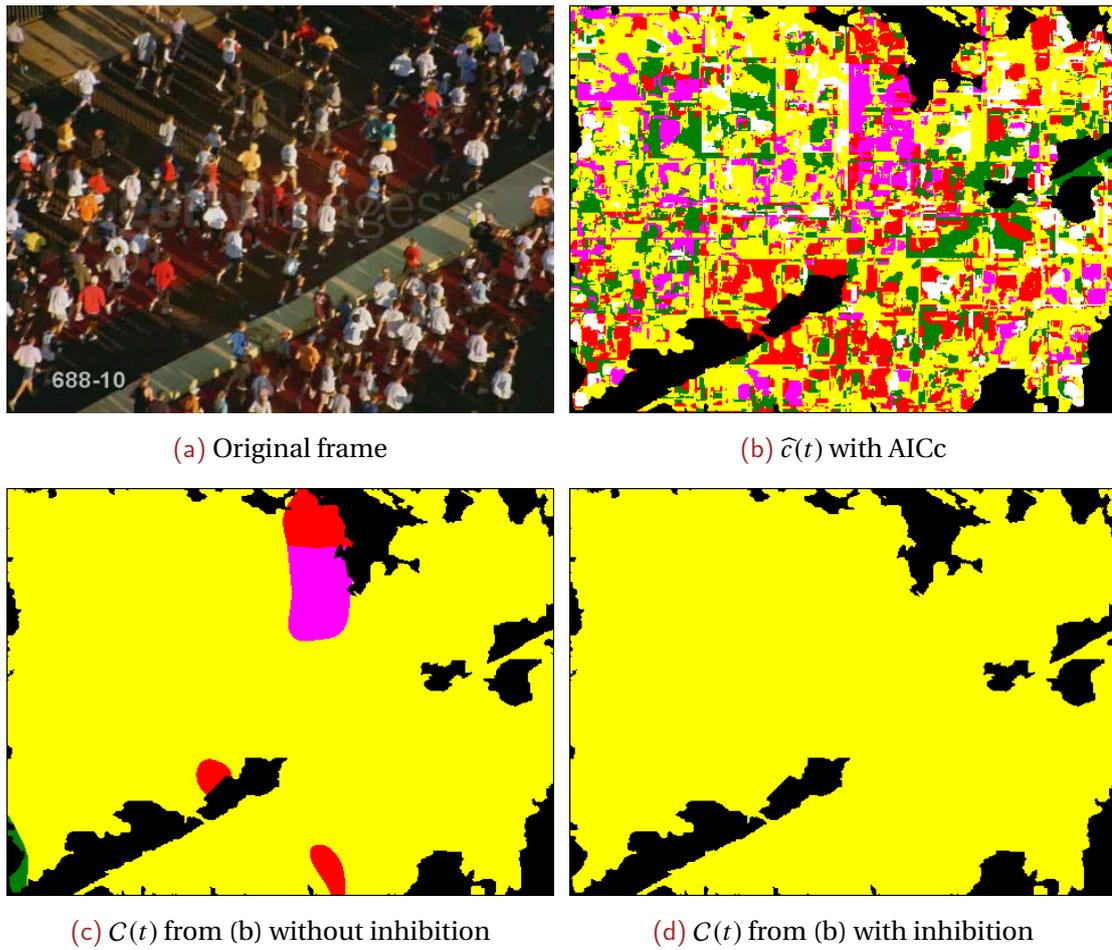


Figure 9.4 – Classification results on the *Marathon lane* sequence where runners move roughly toward the right (○). Class color code is given in Table 9.2.

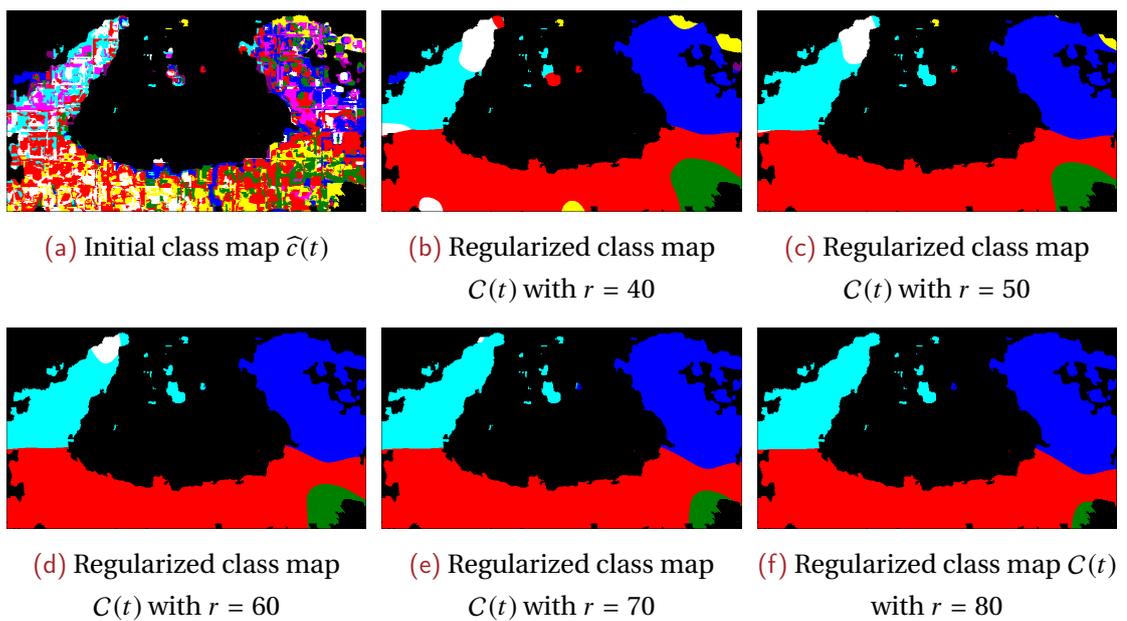


Figure 9.5 – Impact of the radius r of the window function g_r on the regularization

With the lowest values of r , the left branch is segmented in two different motions, due to the perspective effect. Using bigger windows allows us to smooth out this effect and obtain the three main motion patterns, namely Southward translation, counter-clockwise rotation and then Northward translation. Most local patterns, like the small Eastward translations, are also removed. Only a convergence zone remains, which corresponds to shadows in the scene. The latter are visible because of the low local runner density. The convergence class over the runners and their shadow is another perspective effect.

To sum up, the window size affects the classification in the sense that motions smaller than the window size are smoothed out, so that this parameter must be set according to the desired classification “granularity”.

Finally, let us present the classification results of the proposed method for the *Tissue* sequence introduced in Figure 9.1. It shows a group of cells, whose membrane is fluorescently labeled, evolving as a group. The membranes appear in the image as moving edges with similar apparent widths. Since the Laplacian of Gaussian (LoG) filter is also an edge enhancer, we have used our ATLAS method to detect the membranes.

The regularization radius for the classification was set to 50 pixels, which is the scale at which interesting behaviors (e.g., cell divisions) are found. As shown in Figure 9.6, the cells are evolving over time. Globally, the whole cell group is moving toward the lower left corner of the frame. This motion is well recovered as a predominant translation class, divided in both Southward (●) and Westward (●) translations.

From the 60th frame to the end of the sequence, there are large deformations of the cell tissue on top of the tissue translation. Particularly visible is a large stretch of the tissue in the middle of the image (Figure 9.6d,e), which is well recovered as a group divergence (○). Conversely, in the upper part, contracting cells are recovered (●). The large deformation also induces other types of motion around the central stretching.

In addition, smaller events are found over the sequence, such as cell divisions, as shown in Figure 9.6g,h. While not directly detected, due to the very low intensity of the nascent membranes, cell divisions deform the neighboring cells. This results in localized detection of divergence (○) and rotation classes (●).

9.4 Recovery of principal paths

9.4.1 Local path map

In this section, we investigate the recovery of the principal paths followed by the group, with a Eulerian approach. It involves the introduction of local paths determined, in each cell of a grid superimposed on the image, from the space-time average

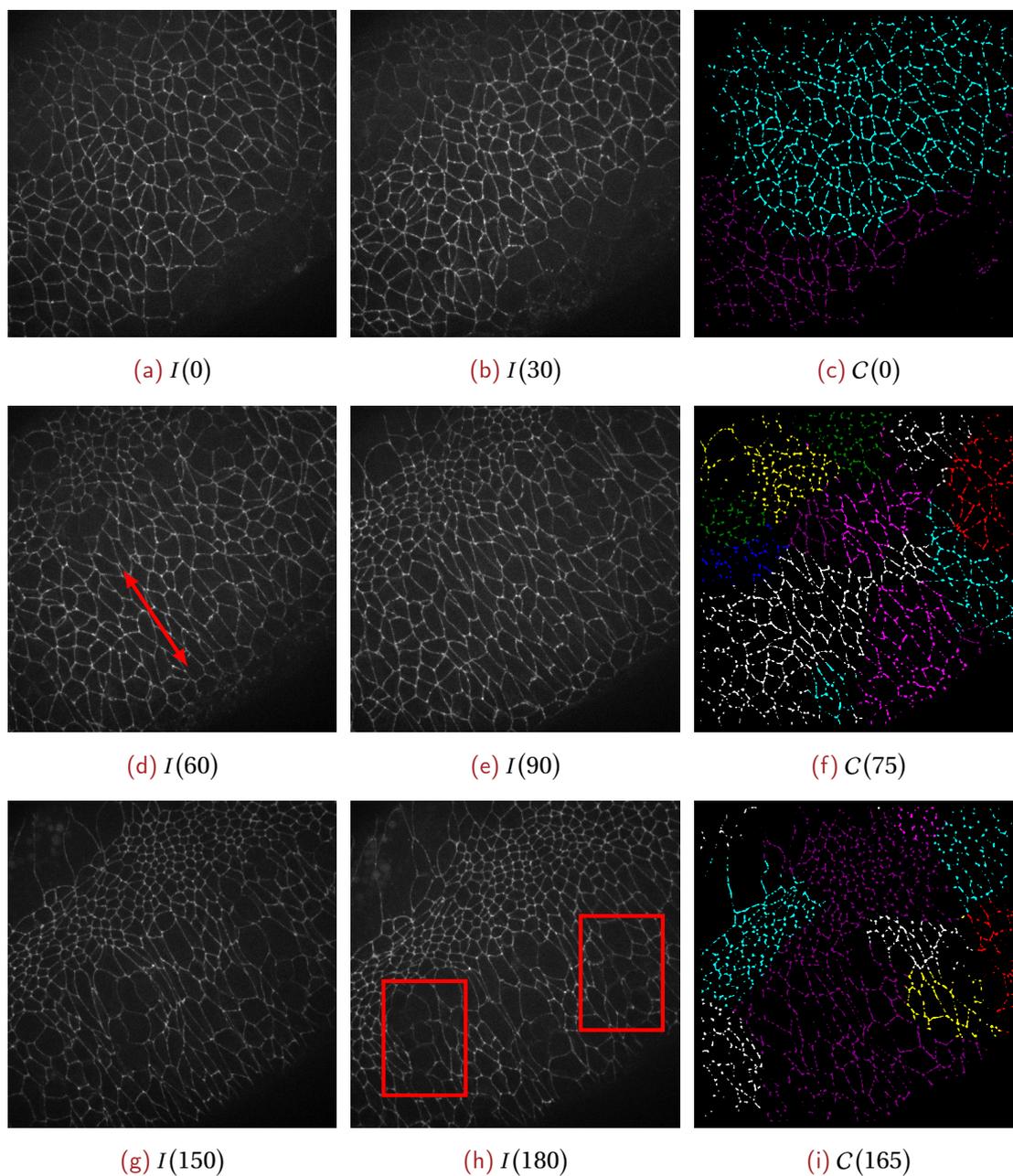


Figure 9.6 – Classification results on the *Tissue* sequence of Figure 9.1. (d) The arrow represents the approximate stretch direction. (h) Regions where cell divisions occur are framed in red. (c,f,i) Class color code is given in Table 9.2.

of the affine flow over five frames. We make the assumption that the motion is stationary, meaning that the paths followed by the group do not vary over the sequence. This is a natural assumption in the frame of the recovery of the principal paths defined as the most followed paths.

Principal paths are simply recovered by accumulating the local paths, and the longest principal path is the dominant one. Thus, as opposed to classical methods, the whole path recovery procedure only relies on local estimates obtained without tracking or advection, and using only five consecutive frames.

Let us first introduce the local path map \mathcal{P} . At each position β of \mathcal{P} will be associated a local path π_i . The set of local paths $\Pi = \{\pi_i\}_{i=1..8}$ corresponds to eight oriented translations: North, Northeast, East, Southeast, South, Southwest, West, Northwest. Indeed, the definition of the π_i 's precisely allows for a 8-connectivity walk in the grid.

The resolution of \mathcal{P} is coarser than the pixelwise resolution of \hat{c} . Since we seek the principal paths in the observed scene, such a spatial integration is required, and it also permits to get more reliable local paths. We generally take blocks β of size 31×31 , but the sensitivity of this parameter is very low, as shown below.

We consider the motion subfields formed at each time instant t by the velocity vectors $\mathbf{w}_{i,\hat{k}}(p, t)$ given by the motion models selected at time t at each point $p \in \beta$ according to equation (9.8). Then, we compute in each block β the mean velocity vector $\bar{\mathbf{w}}(\beta, t)$ in space and time from the motion subfields $\left\{ \mathbf{w}_{i,\hat{k}}(p, u) \middle| p \in \beta, t-2 \leq u \leq t+2 \right\}$ over the video sequence:

$$\bar{\mathbf{w}}(\beta, t) = \sum_{u=t-2}^{t+2} \sum_{p \in \beta} \mathbf{w}_{i,\hat{k}}(p, u). \quad (9.15)$$

The short temporal integration is necessary to ensure that the dominant motion is recovered, as some frames may locally exhibit an ‘‘outlier’’ motion which we do not want to take into account in this section.

Finally, we assign to β the local path label $\pi_i \in \Pi$ which is the closest to the orientation of vector $\bar{\mathbf{w}}(\beta, t)$. Under our stationarity assumption, the local path labels do not depend on time, which will be verified below. Indeed, the eight-quantization of local paths smooths out the slight variations in time of $\bar{\mathbf{w}}(\beta, t)$. The local path then expresses the space-time stationary information on the dominant orientation of the local displacement observed in the block β .

9.4.2 Principal and dominant paths

The principal paths followed by the group in the observed scene are then retrieved from the local path map \mathcal{P} . Starting from one given block in the image, we straightforwardly reconstruct a global path by concatenating the local paths from block to

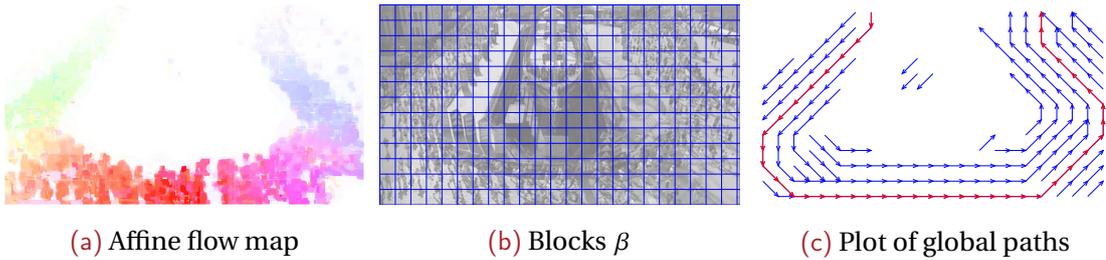


Figure 9.7 – Dominant path recovery on the *Marathon bend* sequence (color code is given in Figure 9.2e). (a) The velocity flow is deduced from the selected motion models. (b) The image is divided into small blocks, where the mean velocity vector in space and time is computed. (c) The global paths are recovered from \mathcal{P}_l and the dominant path is outlined in red.

block following at each step the direction given by each local path among the eight possible ones (N, NW, W, SW, S, SE, E, NE). Then, we cluster all these trajectories and the main clusters supply the principal paths in the observed scene. In particular, the dominant path followed by the group is simply the longest global path.

Let us stress that our approach does not require any tracking of moving points which is a difficult issue in videos of densely crowded scenes. It does not rely on an advection stage from optical flow vectors estimated on a regular grid and on a subsequent trajectory clustering. We can recover the whole set of global paths by starting in turn from every block in the image. Then, the longest path is automatically selected as the dominant one.

The method is illustrated in Figure 9.7.

9.4.3 Experimental results

We have tested our dominant path recovery approach on a wide range of video sequences. Results on a synthetic sequence (*Obstacles*) and several real sequences are presented in Figures 9.7 and 9.8. The accuracy of the detected global path is constrained by the connectivity of \mathcal{P} (8 neighbors) and its resolution, which can be increased by simply reducing the size of the blocks β at the cost of higher computation time.

Figure 9.8 highlights the performance of the method in very different situations. The effect of the block resolution can be observed in the *Marathon lane* sequence, where the true motion direction (East-Northeast) lies between two elements of Π . The processing of the sequence results in the piecewise linear curve for the recovered dominant path (Figure 9.8e), but the dominant path is very close to a straight line for 9×9 blocks (Figure 9.8f). In the other sequences (*Marathon bend*, *Obstacles*, *Tissue* and *Shoal*), the paths are very well recovered, regardless of the block size.

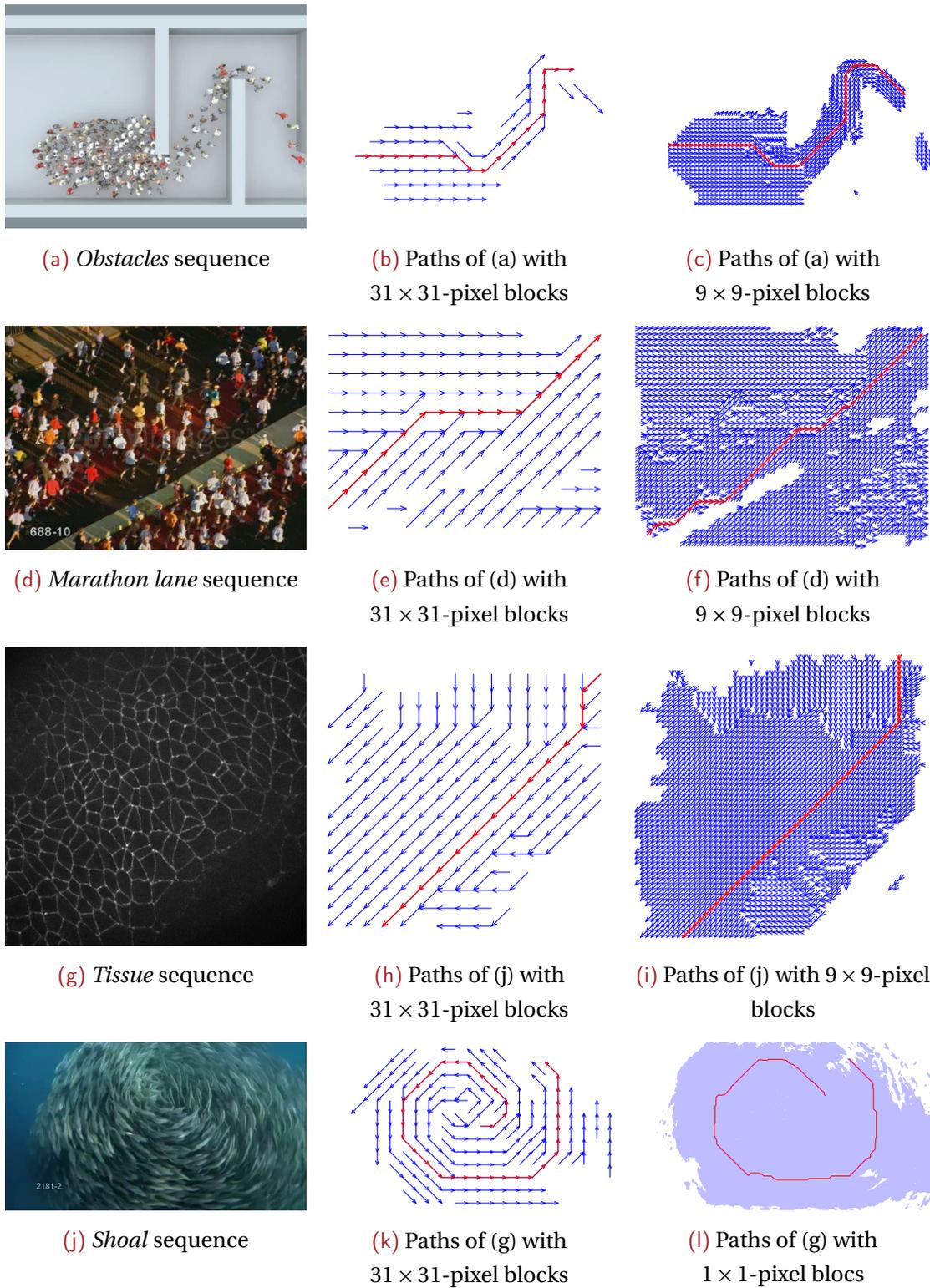


Figure 9.8 – Dominant path recovery (red) in several examples. (a-c) The *Obstacles* sequence, where pedestrians slalom between walls. (d-f) The *Marathon lane* sequence. (g-i) Cells of the *Tissue* sequence. (j-l) Fishes swirling in the *Shoal* sequence.

Computing $\widehat{c}(t)$ and global paths takes about 10 seconds per frame (on the other hand, computing $C(t)$ from $\widehat{c}(t)$ takes less than 1 second) on a 4-core 2.3 GHz laptop, and could be easily parallelized both in space and time.

9.5 Discussion and perspectives

We have proposed a new data-driven overall approach for group motion analysis, which in that context is original both in terms of motion measurements used (LAF, local paths) and in terms of motion classification and path recovery criteria. In contrast to most existing methods which are based on trajectories or tracklets, our approach provides a group motion classification on a frame-by-frame and pixelwise basis. Candidate motion models are estimated in a collection of overlapping windows for each pair of images. An information criterion then allows us to select the best motion candidate in each point of the image domain, allowing to jointly extract the velocity vector and classify the local motion type.

We have then proposed an original method to classify coherent group motions in videos on a frame basis. The group motion classification is achieved with a decision tree regularized with majority votes and involving inhibition between opposed classes. Moreover, since the whole classification process only requires two consecutive frames, even short events can be captured. The algorithm is fast and does not require any learning stage, no fine parameter tuning, and no trajectories computation. The experiments we have carried out demonstrate the accuracy and efficiency of our approach in various real situations.

Finally, we have proposed a method to extract principal paths in a Eulerian way. Local paths are easily obtained by averaging the velocity information of the LAF in small blocks. The principal paths followed by the group can be straightforwardly inferred from the space-time stationary local paths, without any tracking, clustering or learning stage.

Only three parameters are involved in the whole method: the sizes of the windows $\eta(p)$ and $\mathcal{V}(p)$, and of the blocks β . The parameter sensitivity is low in practice, except for the regularization size r , which can be set to attain a user-selected classification granularity.

Future work should include addressing the issue of perspective effects, which can be undesirable in some applications, and focusing on a higher level analysis of the dominant paths.

10

Anomaly detection and localization

IN THIS CHAPTER, we aim at demonstrating the potential of LAF by relying on it for a much more demanding application than the two group motion analyses presented in the previous chapter. Indeed, we target the detection and precise localization of so-called *abnormal events*. The method will be evaluated on videos of crowded scenes which represent very challenging application data, but for which ground-truth can be built by hand. Moreover, apart from the interest we have for videos of crowded scenes for the purpose of method validation, automatic anomaly detection is a much demanded tool in the world of videosurveillance.

Crowd analysis usually requires intense human supervision. When the objective is to analyze anomalies in the scene, sparse events must be identified. This demands an all-time extreme attention from the watcher. Actually, to decide the necessary actions to counteract those anomalies, one has to focus on special events for further analysis, ignoring a vast majority of normal occurrences. This task becomes even more difficult in crowded scenes, where the behavioral complexity in different parts of the video can cause confusion and distraction. Thus, the need for automatic systems that are able to assist the monitoring process of crowded scenes has been growing steadily.

There is no unique definition of an abnormal event. It may depend on the context and the application. As in [Chandola et al., 2009], we consider in this work that anomalies are events that cannot be fit in a proper model obtained from a set of surrounding data. Some samples are given in different situations in Figure 10.1. This formulation is general enough to be of large practical interest. Anomaly is taken here in a broad acceptance of a different behavior with respect to context. It does not mean that the so-called abnormal behavior is necessarily malicious, dangerous, or forbidden. Specifically, we deal with pixel-based action detection and localization, which is finer than frame-based anomaly detection. We want to determine *where* and *when*

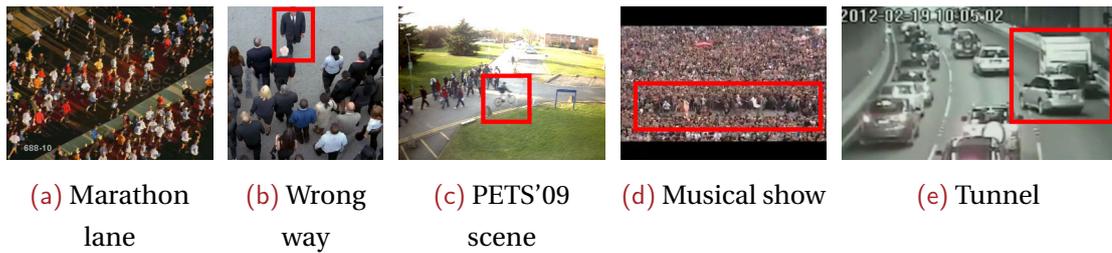


Figure 10.1 – Normal scene and variety of local anomaly cases (framed in red) in group motion scenes. (a) A normal situation where people run together. (b) A man is walking against the crowd; abnormal behavior is localized on him. (c) The cyclist crossing a group of pedestrians is the anomaly. (d) People starting a “circle pit” during a music festival form the abnormal behavior. (e) The car driving into oncoming traffic is the anomaly, along with the one maneuvering to avoid it.

anomalies occur in the successive images of the video sequence, and to be able to do it at the *pixel* level.

The desired solution, however, has to comply with a number of requirements. First, the devised model has to be simple and generic enough so that it can be used in a wide range of applications. Secondly, the algorithm has to be fast. Computational performance is an important criterion looking towards real-time implementation [Lu et al., 2013]. Finally, an anomalous event detection at the frame level does not provide enough information to a human supervisor in order to quickly take actions on the possible anomaly, as pointed out in [Li et al., 2013]. This is even more glaring in densely crowded scenes where it may be utterly difficult to quickly find the localization of subtle anomalies even when it is known beforehand that a given frame contains one. To tackle this problem, the method has to be able to localize anomalies both *temporally* and *spatially*. Specifically, we target the pixelwise segmentation of the anomalies, which has never been done as far as we know.

The common infrastructure for videosurveillance is a network of fixed cameras. Thus, we aim at designing a general efficient solution for anomaly detection and localization in crowded scenes applicable to static cameras. We want it to be instantaneous (or on-line) by delivering frame-by-frame output with a one-frame lag only. Indeed, we use three successive frames to come to a decision at every pixel of every image. We do not formulate any explicit modeling of what a normal situation is supposed to be nor an abnormal one. We are only searching for local discrepancy configuration. As a consequence, we do not need any prior supervised learning. Thus, our method is self-adaptive by being data-driven. It exploits a dense map of local crowd motion classes obtained by maximizing a penalized likelihood criterion and using affine velocity vectors computed in a collection of overlapping windows, as explained in Chapter 9. This chapter presents a novel method for detecting and localizing anomalies in videos of crowded scenes. It is based on the extraction of local histograms of crowd motion

classes over a dense set of patches. These motion class histograms, extracted from LAF, can be seen as relevant motion descriptors. A dedicated distance histogram will be specified and modeled, so that a simple probability of false alarm (PFA) is set by the user to fix the detection sensitivity.

The rest of the chapter is organized as follows. In Section 10.1, we review the related literature and previous work on crowd anomaly detection. The *ad-hoc* histogram distance is proposed in Section 10.2. Then, in Section 10.3, we fully describe our motion-based anomaly detection method and give insights about its main properties. In Section 10.4, we report a comparative and objective evaluation on numerous generated and real images sequences. Finally, we offer concluding comments in Section 10.5.

10.1 Related work

Several approaches have been investigated for anomaly detection in crowd videos. Some methods target specific scenarios, or are specialized for certain types of video data. For instance, escape behaviors can be considered as a specific case of anomaly in surveillance videos [Wu et al., 2014]. However, this is a global type of anomaly since the escape behavior is shared by all the people in the scene from a given starting time. Here, we are interested in local anomaly detection, which means that only a (small) part of the crowd behaves in a different way. Other works are able to detect anomalies locally in videos and without an explicit definition of what the abnormality is. Among these, two main classes are found: trajectory-based methods [Li et al., 2013, Piciarelli et al., 2008, Stauffer and Grimson, 2000, Wu et al., 2010] and feature-based ones [Adam et al., 2008, Antić and Ommer, 2011, Cong et al., 2013, Kim and Grauman, 2009, Kratz and Nishino, 2009, Li et al., 2014].

Trajectory-based methods make use of the relevant information embedded in object tracks [Porikli and Haga, 2004, Stauffer and Grimson, 2000]. Nevertheless, these methods are usually constrained to scenes where it is possible to perform foreground tracking, otherwise they are subject to a large amount of false positives, as pointed out by [Adam et al., 2008]. In [Wu et al., 2010], representative trajectories are first extracted after particle advection and chaotic features are exploited. The normality is modeled by a Gaussian mixture model. A ML estimation with comparison to a predefined threshold enables to determine normal and abnormal frames. Then, anomalies are located within frames identified as abnormal. A different approach was investigated in [Mehran et al., 2009], still based on particle trajectories. Interaction forces between particles are introduced, which yield a force flow in every frame. Recognizing normal frames and abnormal ones in the video sequence is achieved using a bag-of-

words approach involving a latent Dirichlet allocation (LDA) model. Anomalies are delineated in abnormal frames as regions with high force flow.

The method described in [Cui et al., 2011] relied on tracked keypoints to calculate interaction energy potentials, and to separate normal and abnormal crowd behaviors with a support vector machine (SVM) classifier. A non parametric Bayesian framework is designed in [Wang et al., 2011], which can be used to detect anomalous trajectories. Trajectories are described as bags of words, composed of quantized positions and directions. A dual hierarchical Dirichlet process (Dual-HDP) is defined to cluster both words and trajectories. Unlikely trajectories are considered as anomalous ones.

On the other hand, feature-based approaches are less prone to depend on specific scenarios. In [Kratz and Nishino, 2009], spatiotemporal intensity gradients are used, whose distribution over patches in normal situations is supposed to be Gaussian. The Gaussian parameters are learned on the training set. In [Kim and Grauman, 2009], a mixture of probabilistic principal component analysis (MPPCA) aims at modeling normal flow patterns, estimated over patches of the training video set. The method [Chockalingam et al., 2013] relies on probabilistic latent sequential models (PLSM) learned from 15 to 45 minutes of the sequence to analyze, where the user knows that few anomalous events happen. The spatiotemporal compositions (STC) method [Roshtkhari and Levine, 2013] requires about a hundred initialization frames to start learning weights of so-called codewords representing normal behaviors. Afterwards, weights are updated on-line so that no other training sequences are required.

In [Benezeth et al., 2011], co-occurrence matrices for key pixels are embedded in a Markov random field formulation to describe the probability of abnormalities. Mixture of dynamic texture (MDT) are introduced in [Li et al., 2013] with conditional random fields (CRF) to represent crowd behaviors and reported successful results on several datasets, but at the cost of sophisticated models that require intensive learning and high computation time. Other authors focused on giving explicit inclusion of spatial awareness, by subdividing the image in local regions or blocks, in order to obtain a good detection performance with less learning requirements [Adam et al., 2008, Boiman and Irani, 2007].

Another approach was explored in [Antić and Ommer, 2011]. Vectors of spatiotemporal derivatives were utilized as input of a SVM classifier with linear kernel to support the foreground separation process. The latter feeds a graphical probabilistic model. It is interesting to note that such method depends heavily on how well the foreground elements of a video dataset are separated, undermining a possible application for very crowded scenes. Social force models based on optical flow of particles, as introduced in [Mehran et al., 2009] is another example of descriptor used to detect anomalies.

Sparse representations have been increasingly adopted for anomaly detection, as the problem can be elegantly modeled with sparse linear combinations of representations in a training dataset [Cong et al., 2013, Zhao et al., 2011, Zhu et al., 2014]. Explicit image space subdivision can also benefit anomaly localization performance in sparse representation-based methods [Biswas and Babu, 2014]. It is shown in [Mo et al., 2014] that, by introducing nonlinearity into the sparse model, better data separation can be achieved. Also, some modifications can be made to the usual construction of the sparsity models by introducing small-scale least-square optimization steps [Lu et al., 2013], sacrificing accuracy for the benefit of a fast implementation. However, although elegant and sound, sparse representation methods for anomaly detection have not shown high performance in demanding videos.

To summarize, most methods need a large number of frames to perform anomaly detection, which is often stated at the frame level first. They usually require manually labeled sequences or frames to learn models of normal situations, anomalies being defined as outliers of the normal behavior model. In contrast, we aim at detecting and at the same time localizing anomalies directly on a pixelwise basis in every incoming frame following a data-driven “agnostic” approach. We do not build any model of the normal situation and rely on local statistical tests inspired from the center-surround biological vision process [Cavanaugh et al., 2002, Tadin et al., 2006].

10.2 Histogram distance

From the local crowd motion classification maps presented in the previous chapter, we will locate the abnormal behaviors at every pixel p and at every time instant t . As we seek for local anomalies, the classification is not regularized, that is, we rely on \hat{c} instead of C . The detection is based on simple statistics on the $\hat{c}(p, t)$ values, inspired from the center-surround interaction operating in biological visual motion mechanisms [Cavanaugh et al., 2002, Gao et al., 2008, Tadin et al., 2006]. The center-surround interaction has been recently exploited for action recognition [Escobar and Kornprobst, 2012]. It was also influential in the anomaly detection method defined in [Li et al., 2014], but in a different way than ours. We consider spatial surrounding only, while it may be spatiotemporal in [Li et al., 2014] with an extended time interaction. We only evaluate distances between histograms whereas elaborated models are designed (and previously learned) and associated with a Markovian framework in [Li et al., 2014].

For every point p , we compute the occurrences of the eight motion classes of Γ in a patch $\nu(p)$ centered in p , that is, the histogram $\mathcal{H}(p)$ of the $\hat{c}(q, t)$ values for $q \in \nu(p)$. To increase the temporal stability of the method, we compute a spatiotemporal histogram over three consecutive frames so that $\nu(p)$ is a $m \times m \times 3$ space-time patch.

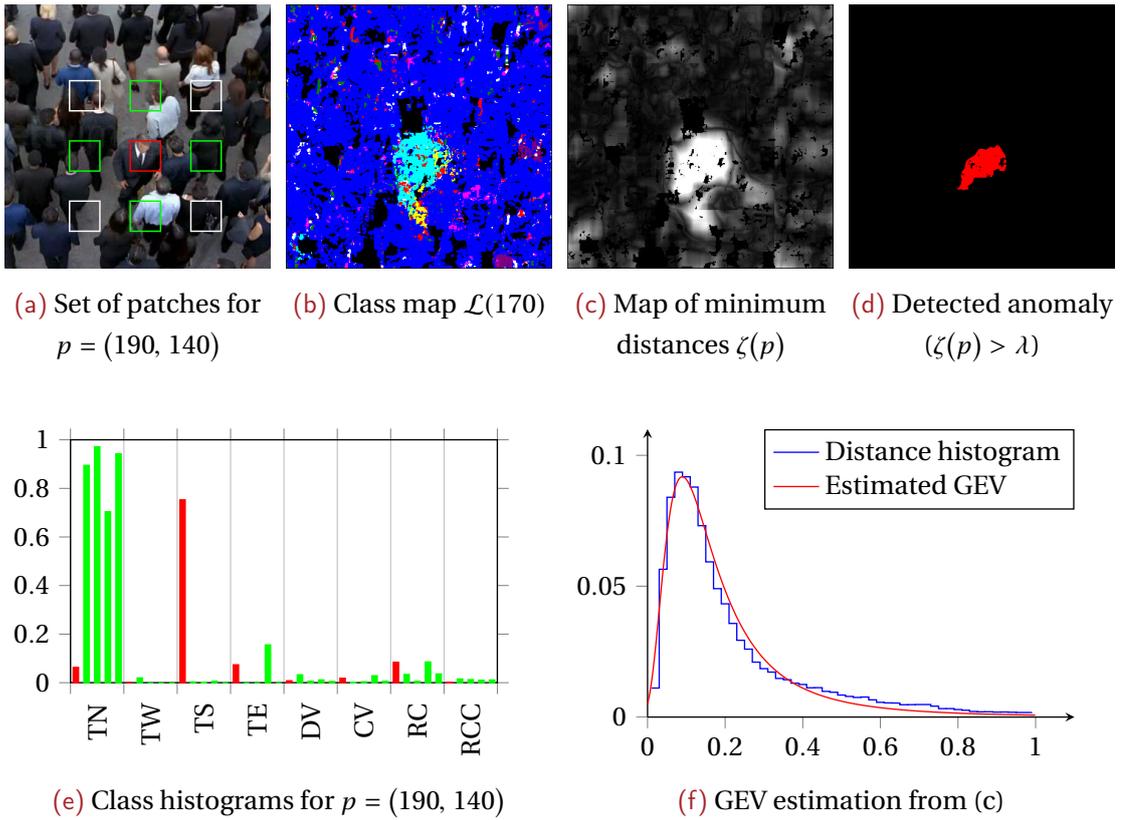


Figure 10.2 – Anomaly detection on the *Wrong way* sequence, where a man is walking to the bottom of the image (south translation \bullet) against a crowd moving in the opposite direction (north translation \bullet). (a) The central patch $v(p)$ is plotted in red at $p = (190, 140)$, along with the neighboring patches $v_i(p)$, $i = 1 \dots 8$. (b) Selected local motion classes at time 170 (class color code of Table 9.2). (c) Map of histogram distances computed at every pixel represented in grey levels (black: $\zeta(p) = 0$; white: $\zeta(p) \geq 1$). (d) Detected anomaly (in red) for a PFA of 1%. (e) Histograms of motion classes for the central (red) and four neighboring patches (green) of 2a. Histogram $\mathcal{H}_c(p)$ (red) is distant enough from the $\mathcal{H}_i(p)$'s to detect an anomaly at p . (f) GEV parameters are estimated from the histogram of distance values of (c).

Then, we take eight neighboring patches $v_i(p)$, $i = 1 \dots 8$, of the same size $m \times m \times 3$, around the patch $v(p)$. Like for constant false alarm rate (CFAR) detection methods [Scharf, 1990], we leave a guarding region around $v(p)$ (of width m), as illustrated in Figure 10.2a.

We compute the histograms $\mathcal{H}_i(p)$, $i = 1 \dots 8$ of the $\widehat{c}(q, t)$ values for $q \in v_i(p)$ in the eight patches $v_i(p)$. Running sums are also used to compute all histograms in constant time, i.e., independently of the patch size. We evaluate the minimum of the distances between histograms $\mathcal{H}_i(p)$ and histogram $\mathcal{H}(p)$:

$$\zeta(p) = \min_{i=1 \dots 8} D(\mathcal{H}_i(p) - \mathcal{H}(p)). \quad (10.1)$$

As a matter of fact, we separate the histograms into two sub-histograms, the first sub-histogram \mathcal{H}^T involving the four translation classes only, the other one $\mathcal{H}^{S,R}$ com-

prising scaling and rotation classes. They are actually of different kind as explained below, and we adopt two different distances for the two categories of sub-histograms.

For the translation class sub-histograms, we resort to the modulo distance introduced in [Cha and Srihari, 2002] for sets of modulo measurements, that is, measurement values forming a ring, which is typically the case for the translation classes and their associated compass directions. It is expressed by:

$$D_{\text{mod}}(\mathcal{H}_i^T(p), \mathcal{H}^T(p)) = \min_{U_i, U} \left(\sum_{l, l'=1}^n d_{\text{mod}}(u_i(l), u(l')) \right), \quad (10.2)$$

where U , resp. U_i , designates the set of translation class values in the patch $v(p)$, resp. in the patch $v_i(p)$. $u(l')$, resp. $u_i(l)$, represents any value of the set U , resp. U_i , once recorded as $\{0, 1, 2, 3\}$ for $\{\text{TN}, \text{TW}, \text{TS}, \text{TE}\}$. n is the number of elements of both U and U_i . In case that the numbers of elements of sets U and U_i are different, a normalization step is added [Cha and Srihari, 2002]. The distance d_{mod} between elements is given by:

$$d_{\text{mod}}(u_i(l), u(l')) = \begin{cases} |u_i(l) - u(l')| & \text{if } |u_i(l) - u(l')| \leq \frac{|\Gamma_T|}{2}, \\ |\Gamma_T| - |u_i(l) - u(l')| & \text{otherwise.} \end{cases} \quad (10.3)$$

where Γ_T is the subset of the four translation classes, and $|\Gamma_T| = 4$. Since the distance between two histograms can be expressed in terms of the distances of element measurement values, criterion (10.2) allows us to find the minimum difference of pair assignments between sets U and U_i .

The interest of the modulo distance is that the distance between opposite translation directions (e.g., North and South directions) will be higher than the one between adjacent translation directions (e.g., North and West directions). Such a behavior is recommendable in our case, all the more so as we deal with a coarse direction quantization. Indeed, points undergoing a given slanted translation (e.g., North-West) could be shared out among two adjacent translation classes (North and West translation classes to continue our example).

After testing several usual histogram distances, we have retained the L_1 distance for comparing histograms $\mathcal{H}_i^{\text{S,R}}$ and $\mathcal{H}^{\text{S,R}}$:

$$D_{L_1}(\mathcal{H}_i^{\text{S,R}}(p), \mathcal{H}^{\text{S,R}}(p)) = \sum_{c \in \Gamma_{\text{S,R}}} \left| \mathcal{H}_i^{\text{S,R}}(p, c) - \mathcal{H}^{\text{S,R}}(p, c) \right|, \quad (10.4)$$

where $\Gamma_{\text{S,R}} = \{\text{SN}, \text{SP}, \text{RN}, \text{RP}\}$ is the subset of the four scaling and rotation motion classes, and $|\Gamma_{\text{S,R}}| = 4$.

Finally, we have:

$$D(\mathcal{H}_i(p) - \mathcal{H}(p)) = D_{\text{mod}}(\mathcal{H}_i^T(p), \mathcal{H}^T(p)) + D_{L_1}(\mathcal{H}_i^{\text{S,R}}(p), \mathcal{H}^{\text{S,R}}(p)), \quad (10.5)$$

with equally weighted distances, since the ranges of the modulo and L_1 distances are similar as explained in [Cha and Srihari, 2002].

10.3 Pixel-level anomaly detection and localization

Since we have defined an abnormal behavior as revealed by an irregular motion pattern with respect to the surrounding motion patterns, an anomaly will be detected at point p if and only if:

$$\zeta(p) > \lambda, \quad (10.6)$$

with variable $\zeta(p)$ defined in (10.1). The minimum distance $\zeta(p)$ for all the points $p \in \Omega_7^*$ follows a generalized extreme value (GEV) distribution [Embrechts et al., 1997], which is defined by:

$$F(\zeta(p), \alpha, \beta, \kappa) = \exp \left[- \left(1 - \kappa \frac{\zeta(p) - \beta}{\alpha} \right)^{1/\kappa} \right], \quad (10.7)$$

where α , β and κ are respectively the width, location and shape parameters of the GEV distribution. First, we have to estimate its parameters. We have adopted the mixed L-moments/maximum likelihood method described in [Boulangier et al., 2010a]. An example is given in Figure 10.2f. Then, we set a PFA to automatically set the threshold value λ as a quantile of the GEV distribution, in order to detect the local abnormal behaviors in the observed crowd. In practice, PFA will be set between 0.5 and 2%. As explained later, this value can be seen as the expected false positive rate (FPR), so the detection sensitivity can easily be set according to the targeted application.

10.4 Experimental results

To demonstrate the performance of our method, we need to show that it detects anomalies in abnormal situations, the anomalies are well segmented, and it does not detect anomalies in normal situations.

10.4.1 Simulated cases

We first deal with simulated cases to make an easy objective evaluation of different aspects of our method. We have taken a sequence depicting a normal behavior of a dense moving crowd (Figure 10.3a), and we have added a simulated anomaly (Figure 10.3b). Specifically, we have cropped a rectangular small region from the synthetic sequence *Corridor* where pedestrians are walking from left to right. After mirroring it, we have inserted it again in the original sequence but with a translation from right to left. Thus, one pedestrian is walking in the opposite direction of the whole group and constitutes an anomaly in the resulting *Modified corridor* sequence. The anomaly is correctly detected (Figure 10.3c). Conversely, very few false detections occur in the original *Corridor* sequence without anomaly. More precisely, as explained later, the FPR is indeed really close to the user-selected PFA.

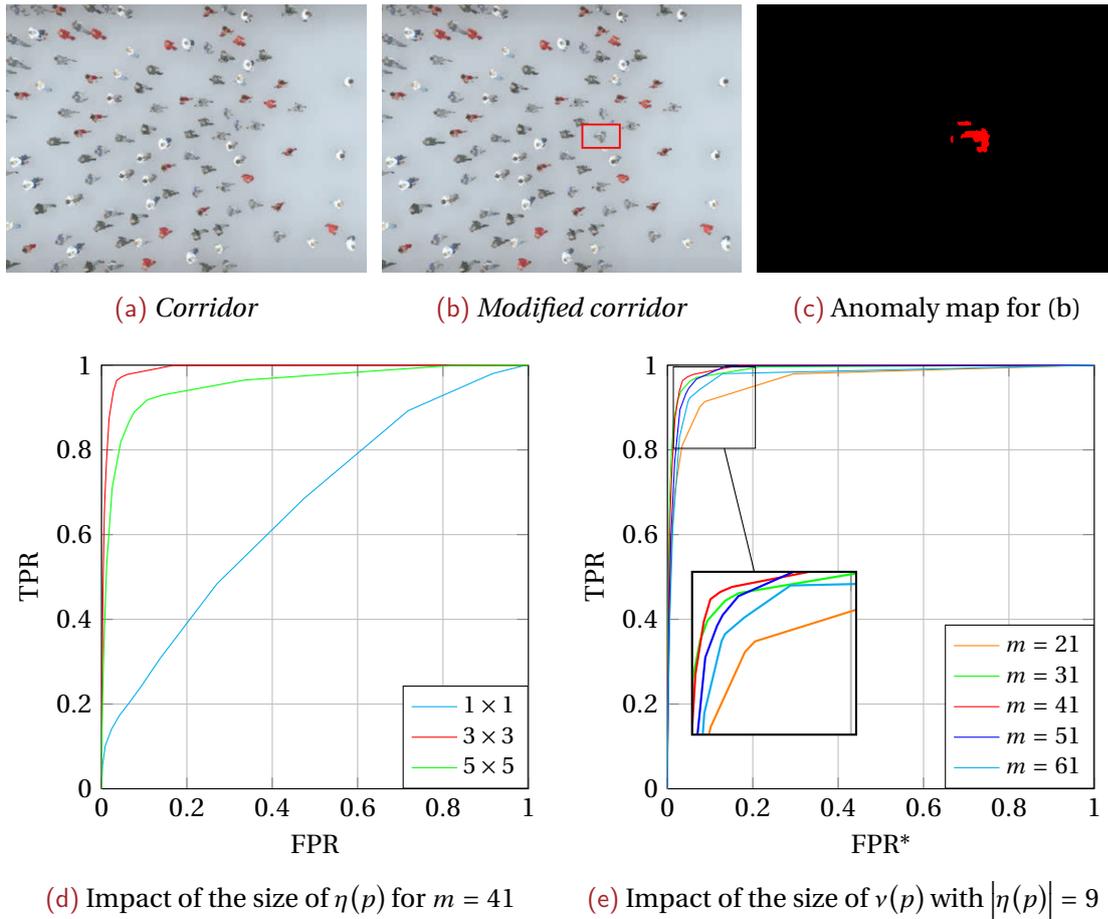


Figure 10.3 – Anomaly detection results and impact of the size of $\eta(p)$ and the size ($m \times m \times 3$) of $\nu(p)$ for the *Modified corridor* sequence. (a) Original *Corridor* sequence. (b) *Modified corridor*, where the anomalous region is framed in red. (c) Anomaly map for PFA = 1%, $m = 41$ and $|\eta(p)| = 9$. (d,e) We vary the PFA value to plot the ROC curves (zooming is performed on a portion of the curves). Best results are obtained with $m = 41$ and $|\eta(p)| = 9$, and, in this case, the area under the curve is equal to 0.99.

We have computed receiver-operator characteristic (ROC) curves, that is, the evolution of true positive rate (TPR) versus FPR, for different sizes of local neighborhood ν_p used to compute the AICc, and different sizes of patches $\nu(p)$ used to compute local histograms, as reported in Figure 10.3d,e. Since our method supplies pixelwise anomaly localization, TPR is the proportion of anomaly pixels that are correctly detected, and FPR is the proportion of normal points that are detected as anomalies. The optimal size for the affine model selection neighborhood ν_p is 3×3 . The same size will be adopted for all the processed sequences. Concerning square patches $\nu(p)$, best results are obtained with patch width between 31 and 51 pixels, which is close to the cropped region size. For this patch width range, the area under ROC curve is equal to 0.99.

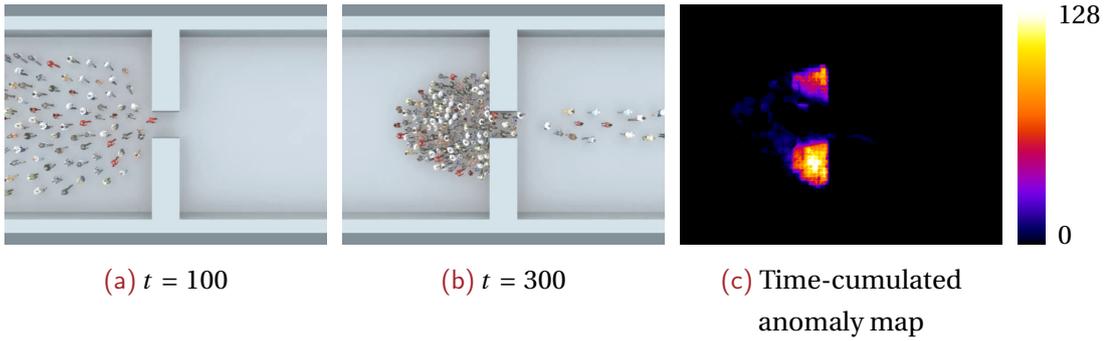


Figure 10.4 – Temporal analysis of anomaly localization over the synthetic *Escape* sequence. (a–b) People are leaving the room toward the right through a single door. (c) Map of the number of times an anomaly is detected at each point over 400 frames with PFA = 2% (color scale is given on the right).

The temporal stability of our method is illustrated in Figure 10.4. In the *Escape* sequence, individuals leave a room by a single door, resulting in a local congestion, a potentially dangerous situation which must be classified as an abnormal behavior. Since the location of the abnormal behavior regions are stationary throughout the sequence, we have accumulated the successive binary detection maps. As expected, the most frequent detections clearly lie on both sides of the door (Figure 10.4f) where people are stuck.

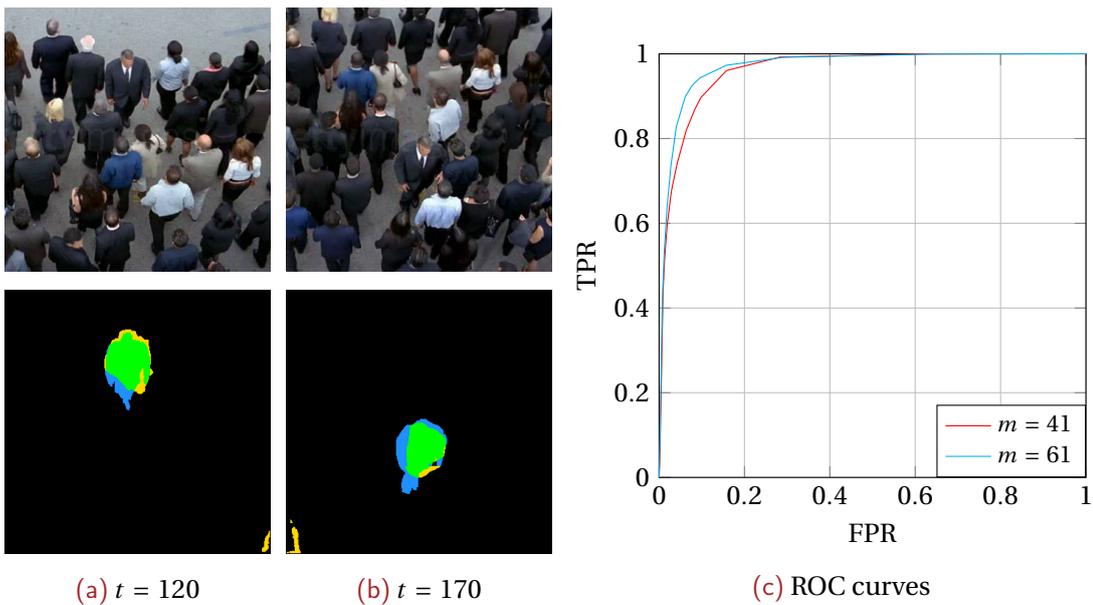


Figure 10.5 – Results on the *Wrong way* sequence. (a,b) Detection results on the first and last labeled frames, with $m = 61$ and PFA = 1% (green: TP; blue: FN; yellow: FP). (c) ROC curves for $m = 41$ and $m = 61$. Area under curve is equal to 0.97 for $m = 61$.

10.4.2 Real scenes

To the best of our knowledge, there is no benchmark available for pixelwise anomaly localization in dense crowds, and such an evaluation has not been attempted so far because most existing methods are not designed for both detecting and localizing anomalies on a pixel basis. The UMN dataset and the web dataset of [Mehran et al., 2009] are concerned with globally abnormal crowd behaviors (as escape panics). The UCSD dataset [Li et al., 2014] mainly involves very sparsely crowded scenes. Besides, the pixel-level evaluation used in [Li et al., 2014] remains a frame-based evaluation taking into account a sufficient recall (40%) of abnormal regions defined by bounding boxes in the ground truth. We propose a more demanding evaluation criterion: we want to carry out a *pixel-based* objective evaluation. To that end, we have manually determined the pixelwise anomaly ground truth for two real sequences: the *Wrong way* sequence and one sequence from the PETS dataset. Thus, we are able to supply a truly pixel-based objective evaluation for real anomaly cases.

In Figure 10.5, we report results of our anomaly localization method on the *Wrong way* sequence (Figure 10.1b). In this crowd scene, one pedestrian is walking against the crowd. As shown at four different time instants, respectively in Figure 10.1b, Figure 10.2d, and Figure 10.5a,b, the man heading to the bottom of the frame is well segmented. ROC curves established for the *Wrong way* video are very good. Area under the curve for $m = 61$ is equal to 0.97 and equal to 0.96 for $m = 41$. The camera being closer to the crowd than in previously processed sequences, pedestrians look bigger. The bounding box of the man walking downward has a width of about 60 pixels all over the sequence, while its height in the image varies between 50 and 100 pixels due to partial occlusions. The best detection rate is obtained with square patches of 61-pixel width. It is not possible to compare with [Mehran et al., 2009] since the authors have only provided ROC curves for abnormal frame detection and color-coded force flow maps, they did not supply any binary detection maps.

In Figure 10.6, we demonstrate the robustness of our method on a sequence of the PETS'09 dataset¹. In this sequence, we aim at detecting the only pedestrian walking toward the left of the frame and passing *behind* a group of people going in the opposite direction. The task is particularly challenging here, because the group often occludes the lonely pedestrian and the pedestrian is very small in the image. Yet, the lonely pedestrian is accurately segmented and no false alarm occurs as displayed in Figure 10.6a,b. The area under ROC curve is 0.93 with a patch size of 21 pixels, similar results were obtained for a size of 31 pixels (Figure 10.6c).

Supplementary results obtained on other real sequences involving various behaviors (presented in Figure 10.1) are reported in Figure 10.7. They confirm the ability of

¹<http://www.cvg.rdg.ac.uk/PETS2009>

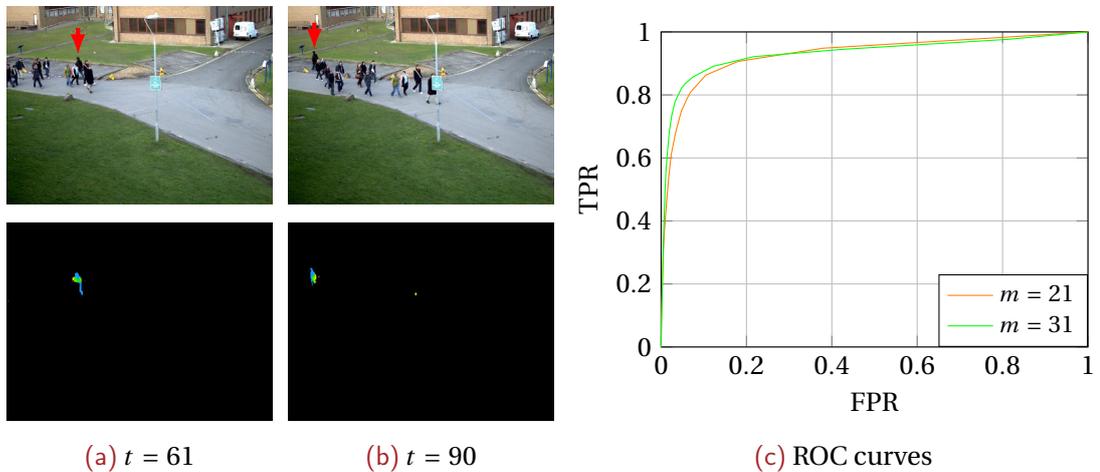


Figure 10.6 – Results on a sequence of the PETS'09 dataset. (a,b) Top row: Input frames (the red arrow points to the pedestrian to be detected). Bottom row: Anomaly maps with $m = 31$ and PFA = 1% (green: TP; blue: FN; yellow: FP). (c) ROC curves for $m = 21$ and $m = 31$ (area under curve is equal to 0.93 for both).

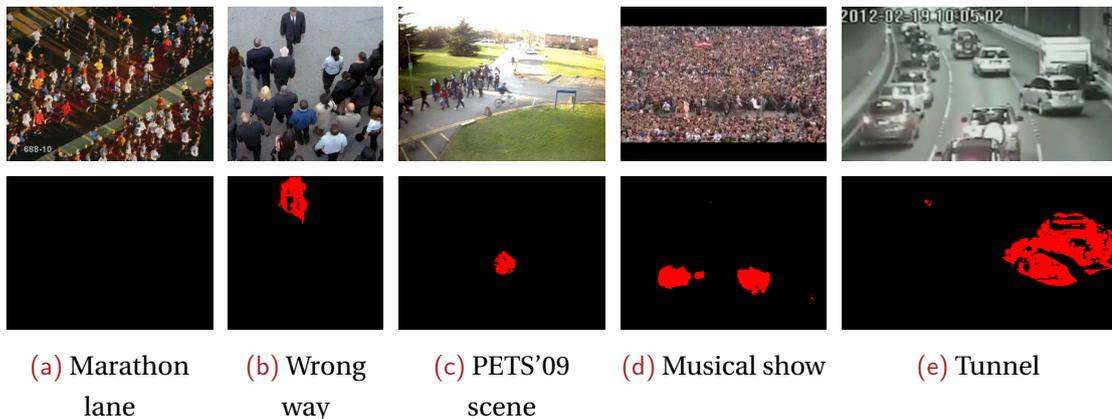


Figure 10.7 – Anomaly detection for the samples presented in Figure 10.1. Top row: input images. Bottom row: Anomaly maps computed with our method with the same parameter setting (PFA = 1% and $m = 61$) apart from (e) where $m = 81$. (a) No anomaly is detected in this normal situation. (b) The man who is walking against the crowd, (c) the cyclist crossing a group of pedestrians, (d) people running in the crowd, and (e) the car driving into oncoming traffic and the one maneuvering to avoid it are correctly detected.

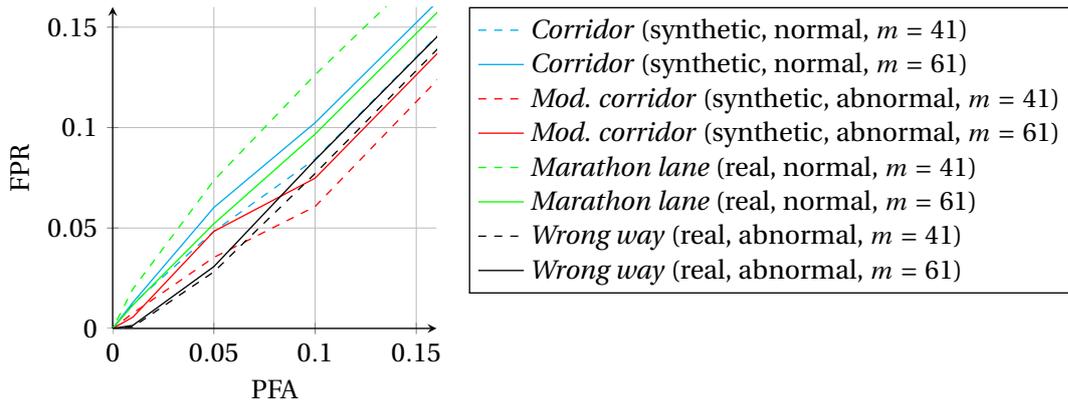


Figure 10.8 – Plots of the curves expressing the relationship between PFA and FPR for various synthetic and real sequences with or without anomalies.

our method to reliably segment regions with abnormal behaviors while not detecting anomalies in normal situations.

10.4.3 FPR control, parameter setting and computation time

Figure 10.8 contains plots expressing the relationship between PFA and FPR for different sequences (synthetic and real ones, without and with anomalies) and for two different sizes of patches $\nu(p)$. *Corridor* and *Modified corridor* were introduced in Figure 10.3a,b. In the *Marathon lane* sequence, people run from left to right, and nothing anomalous happens (one image of the sequence is given in Figure 10.1a). As expected, no anomalies were detected as shown in Figure 10.1a (apart from a few spurious points isolated in space and time). Finally, the *Wrong way* sequence was presented in Figure 10.5. In every case, the obtained FPR is close to the PFA parameter value since each plot in Figure 10.8 is close to the straight line of slope 1. The same holds for all the other sequences we have analyzed. Thus, a clear advantage of the proposed method is that the FPR can be easily controlled by the user through the setting of PFA.

10.4.4 Parameter setting and computation time

Overall, the anomaly localization is accurate and stable over time. The method involves two parameters only, easy to set: PFA value which can be seen as an expected FPR, and dimensions of ν which must be close to those of the expected anomaly bounding box – typically the size of the image projection of a pedestrian. It can be deduced from the scene characteristics and the camera setting. Furthermore, as reported in Figure 10.2e, Figure 10.5c and Figure 10.6c, the sensitivity of the patch size is low, so that it does not need to be finely tuned.

The computation time for the largest sequences we have processed (of image size 720×576) is of 6 to 10 seconds per frame on a laptop with a 4-core 2.3 GHz processor. Let us notice that using running sums, the computation load of the criterion AICc remains low and does not depend on the patch size. Even if our method is not real-time so far, we have described it as on-line, since by construction it delivers the anomaly map at each time instant with a one-frame lag only. It can be highly parallelized to eventually attain real-time processing. On-line detection is unreachable with approaches somehow relying on temporal integration, regardless of the processing power.

10.5 Discussion

We have developed an original data-driven method for anomaly detection in videos of dense crowds. It yields a pixelwise anomaly localization map at every time instant using only three successive frames. The method relies on the computation of parametric motion models, on the local selection of the crowd motion classes with the AICc penalized likelihood criterion, and on the bio-inspired evaluation, at every pixel, of a distance between center-surround histograms of the local crowd motion classes. The method is simple to perform, there is no normal or abnormal behavior models to learn and it is almost parameter-free. It can handle a large category of situations. We have reported experimental results on both simulated and real cases. The objective evaluation has demonstrated the accuracy and reliability of the method.

Further work should investigate the impact of different geometric center-surround configuration of patches $\nu(p)$ and the weighting of the class histograms by the velocity vector magnitude, to fully take advantage of the LAF data. The latter could enable to discriminate not only with respect to the motion class, but also between different speeds in a same class.

11

General conclusion

THE THESIS aimed at detecting, estimating and characterizing dynamical events in image sequences. We have focused on two types of dynamics, namely individual and group dynamics. In each case, we have proposed automatic methods to calibrate algorithm parameters by using statistical arguments. Local approaches allowed us to design computationally light and efficient methods able to cope with complex image sequences. We have mainly been concerned with fluorescence microscopy image sequences and videos of crowded scenes.

Within the scope of fluorescence microscopy, we have principally focused on the exocytosis, a dynamical process by which some molecules are transported from the cell interior to the extracellular medium. Specifically, we have studied the behavior of three proteins, Transferrin receptor (TfR), Langerin and Rab11, during the late steps of exocytosis. This includes transporting the molecules and passing through the cell frontier (plasma membrane). Molecule transport is mediated by the so-called vesicles, which enclose the molecules and then literally fuse to the plasma membrane to free molecules outside the cell.

By using total internal reflection fluorescence microscopy (TIRFM), we have studied in particular the vesicle fusion dynamics. In this analysis workflow, vesicles and fusion events, appearing as bright spots in the images, must first be detected. Therefore, we have developed a new spot detection method ATLAS based on adaptive thresholding of LoG images with auto-selected scale. After detecting the fusion events, we have proposed models and methods to estimate the motion of the proteins at the vesicle fusion.

We have finally investigated methods for the study of dynamics involving interactions between individuals. We have introduced a motion measurement data, labeled affine flow (LAF), which involves both the pointwise velocity vectors and the local motion type, using only two consecutive frames. This hybrid data allowed us to target

various applications, including the characterization of the motion of the group as a whole, the recovery of main paths in the scene, and the detection and localization of abnormal behaviors in the group.

11.1 Main contributions and results

11.1.1 Spot detection

First, we have investigated methods for the detection of spots in TIRFM images. In such images sequences, vesicles appear as small bright spots of similar size. Therefore, our spot detection method relies on the similar size hypothesis as a prior knowledge to improve the detection performance. Considering the Laplacian of Gaussian (LoG) to enhance the spots, the best detection scores are obtained if the LoG scale is accurately selected beforehand. The optimal filter scale is automatically selected in a scale-space representation of the image. The spot detection then amounts to a pointwise thresholding of the LoG image computed at this scale, and the threshold adapts to the local image statistics. While designed in view of the exocytosis detection in TIRFM, the proposed adaptive thresholding of LoG images with auto-selected scale (ATLAS) method can be applied to a wide variety of images and delivers a leading performance.

11.1.2 Individual dynamics analysis

In the frame of the exocytosis study, there are two cases depending on the tagged protein. As for Rab11, it undergoes a translational motion before the vesicle fusion to plasma membrane, and then diffuses. Therefore, we have proposed a classification method to distinguish between the two dynamic situations. On the other hand, when observing a transmembrane protein tagged with pHluorin, vesicle fusion to plasma membrane is visible in the image sequences as spot appearances. In this case, we simply applied ATLAS to the sequence of temporal difference images to detect these spots.

Given the detected space-time fusion events, the biological models for the proteins diffusion can be explored. In particular, we have proposed a new fusion model, termed “small-extent source with exponential decay release” (SSED). As opposed to the existing point source model for which all the contents of the vesicle start diffusing at a given time t_0 , the vesicle is supposed to slowly release the proteins after fusion in the SSED model. This motivates the introduction of another biophysical parameter, the residence time, in addition to the diffusion coefficient.

Furthermore, the new SSED vesicle fusion model is then estimated with a method which proved to perform better than existing estimation methods on the simpler

point source model. Dynamical behaviors have been characterized in a set of image sequences, with the use of the SSED model, to exhibit the different behaviors of the two transmembrane proteins TfR and Langerin.

11.1.3 Group dynamics analysis

In the last part of the thesis, we have focused on the development of LAF, a new motion measurement, which is more appropriate to group analysis than classical detect-and-track methods, in the sense that it captures both the type and the numerical value of the local velocities without requiring any segmentation. Thus, LAF combines the advantage of quantitative and qualitative motion descriptors.

As an application, we have shown the relevance of this motion descriptor to characterize group motion by classifying motion patterns and by recovering the principal paths followed by the group. We have finally proposed a new method to detect and segment local anomalous behaviors, still building on LAF. All these methods only rely on very short temporal integration, and are therefore suitable for on-line processing, as opposed to most existing group motion analysis methods.

11.2 Ongoing and future work

We are currently working on a multiscale extension of our spot detector, relying on the same scale-space representation, but with a more elaborated selection criterion. Preliminary results were presented in Part I, but a more advanced method is envisaged. It relies on a probabilistic formulation of the selection criterion, which allows us to set a meaningfulness threshold. This way, the user does not need to specify the number of scales. Preliminary results on the scale set selection itself are already promising, but the adaptive thresholding must be extended to cope with multiple scales. Then, detection performance must be evaluated to verify that the automatic scale set selection behaves better than multiscale detectors like the multiscale variance stabilizing transform (MS-VST).

The proposed SSED model will be used in other experimental conditions to investigate possible differences between mutant and wild cells. In particular, the team of Jean Salamero at Institut Curie aims at investigating the role and interactions of Actin in the recycling process. Ultimately, the dynamics of Rab11 are targeted, since open questions remain for this protein.

This leads us to the future work to be accomplished regarding Rab11. The behavior of Rab11 should be modeled and estimated thoroughly in 3D+time image sequences acquired with multi-angle TIRFM [Boulanger et al., 2014]. At this time, preliminary results and discussions exposed in this thesis will constitute some basis for the development of 3D+time estimation methods.

It is already known that Rab11 finally returns to the cytosol at a given, but unknown, time. Understanding the timeline of the membrane fusion process, and in particular the position in time of Rab11's dissociation, is of primary interest for cell biology. Dissociation should therefore be modeled and estimated in 3D+time image sequences, constituting the short-term target of the study. In particular, the different Rab11 dynamical models presented in Part II should be challenged. Then, efforts should be put in the identification of interactions between proteins, in particular Rab11 with TfR and Rab11 with Langerin, in order to point out potential behavioral differences.

At a low level modeling, other assumptions could be relaxed, for example the membrane isotropy, homogeneity and planeity. These improvements would naturally impact both the models and estimation procedures, and open up a large area of possible studies to end up with a fine model, and reliable estimation and classification methods.

As for group motion analysis, we have presented interesting results for crowd motion analysis throughout Part III, and shown some qualitative examples of cell group dynamics.

In video analysis, working on anomaly detection should be further investigated, for example with the use of the magnitude of the velocity vectors in the histogram descriptors or with another decision paradigm. This is currently investigated with Juan-Manuel Pérez-Rúa, a PhD student in the Serpico team.

In addition, the group motion study is still a proof of concept which we think could find a vast set of biological application domains. The agnostic approach we have proposed is interesting from an image processing point of view, but better results would probably be obtained by adding specific priors to cope with a particular issue. Anomaly detection is a classical topic in biology, found for example in tissue genesis.

As pointed out by a reviewer of this thesis, an interesting application could be active polar gels, where biological knowledge is available to enrich the affine models and therefore improve the biological relevance of the method. Given the immensity of the biological research landscape, we are confident that other applications could find their way in PhD theses to come.

Bibliography

- [Abramowitz and Stegun, 1972] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Number 55. Courier Dover Publications.
- [Adam et al., 2008] Adam, A., Rivlin, E., Shimshoni, I., and Reinitz, D. (2008). Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(3):555–560.
- [Aguet, 2009] Aguet, F. (2009). *Super-resolution fluorescence microscopy based on physical models*. PhD thesis, École polytechnique fédérale de Lausanne.
- [Airy, 1835] Airy, G. B. (1835). On the Diffraction of an Object-glass with Circular Aperture. *Trans. Cambridge Philosophica*, 5:283–291.
- [Almeida and Vaz, 1995] Almeida, P. F. F. and Vaz, W. L. C. (1995). Lateral diffusion in membranes. *Handbook of Biological Physics*, 1:305–357.
- [Antić and Ommer, 2011] Antić, B. and Ommer, B. (2011). Video parsing for abnormality detection. In *2001 IEEE Conf. Computer Vision (ICCV)*, pages 2415–2422.
- [Avery et al., 1944] Avery, T. E. X., MacLeod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribo-nucleic acid fraction isolated from pneumococcus type III. *J. Experimental Medicine*, (79):137–158.
- [Axelrod, 2008] Axelrod, D. (2008). Total internal reflection fluorescence microscopy. *Methods in Cell Biology*, 89:169–221.
- [Axelrod et al., 1976] Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E., and Webb, W. W. (1976). Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophysical J.*, 16(9):1055–1069.
- [Barlow and Guerin, 2007] Barlow, A. L. and Guerin, C. J. (2007). Quantization of widefield fluorescence images using structured illumination and image analysis software. *Microscopy Research and Technique*, 70(1):76–84.

- [Basharat et al., 2008] Basharat, A., Gritai, A., and Shah, M. (2008). Learning object motion patterns for anomaly detection and improved object detection. In *2008 IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8.
- [Basset et al., 2014a] Basset, A., Boulanger, J., Bouthemy, P., Kervrann, C., and Salamero, J. (2014a). SLT-LoG: A vesicle segmentation method with automatic scale selection and local thresholding applied to TIRF microscopy. In *2014 IEEE Int. Symp. Biomedical Imaging*, Beijing.
- [Basset et al., 2014b] Basset, A., Bouthemy, P., Boulanger, J., Salamero, J., and Kervrann, C. (2014b). Detection and classification of dynamic subcellular events in TIRF microscopy sequences. In *2014 IEEE Int. Symp. Biomedical Imaging*, Beijing.
- [Basset et al., 2013] Basset, A., Bouthemy, P., and Kervrann, C. (2013). Frame-by-frame crowd motion classification from affine motion models. In *10th IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, Krakow.
- [Baumgartner et al., 2013] Baumgartner, T., Mitzel, D., and Leibe, B. (2013). Tracking people and their objects. In *IEEE Conf. Computer Vision and Pattern Recognition*, Portland.
- [Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522.
- [Benezeth et al., 2011] Benezeth, Y., Jodoin, P.-M., and Saligrama, V. (2011). Abnormality detection using low-level co-occurring events. *Pattern Recognition Letters*, 32:423–431.
- [Biswas and Babu, 2014] Biswas, S. and Babu, R. V. (2014). Sparse representation based anomaly detection with enhanced local dictionaries. In *2014 IEEE Conf. Image Processing*, pages 5532–5536.
- [Boiman and Irani, 2007] Boiman, O. and Irani, M. (2007). Detecting irregularities in images and video. *Int. J. Computer Vision*, 74(1):17–31.
- [Boulanger, 2007] Boulanger, J. (2007). *Non-parametric estimation and contributions to image sequence analysis: Modeling, simulation and estimation of the intracellular traffic in video-microscopy image sequences*. PhD thesis, Université de Rennes 1.
- [Boulanger et al., 2010a] Boulanger, J., Gidon, A., Kervrann, C., and Salamero, J. (2010a). A patch-based method for repetitive and transient event detection in fluorescence imaging. *PLoS One*, 5(10):e13190.

- [Boulanger et al., 2014] Boulanger, J., Gueudry, C., Münch, D., Cinquin, B., Paul-Gilloteaux, P., Bardin, S., Guérin, C., Senger, F., Blanchoin, L., and Salamero, J. (2014). Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging. *Proc. National Academy of Sciences*, 111(48):17164–17169.
- [Boulanger et al., 2009] Boulanger, J., Kervrann, C., and Bouthemy, P. (2009). A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery. *Medical Image Analysis*, 13(1):132–142.
- [Boulanger et al., 2010b] Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J.-B., and Salamero, J. (2010b). Patch-based non-local functional for denoising fluorescence microscopy image sequences. *IEEE Trans. Medical Imaging*, 29(2):442–453.
- [Breen et al., 1991] Breen, E. J., Joss, G. H., and Williams, K. L. (1991). Locating objects of interest within biological images: The top hat box filter. *Computer Assisted Microscopy*, 3(2):97–102.
- [Bright and Steel, 1987] Bright, D. S. and Steel, E. B. (1987). Two-dimensional top hat filter for extracting spots and spheres from digital images. *J. Microscopy*, 146(2):191–200.
- [Brown, 2006] Brown, D. (2006). Imaging protein trafficking. *Nephron Experimental Nephrology*, 103(2):e55–e61.
- [Brown, 1828] Brown, R. (1828). A brief account of microscopical observations made on the particles contained in the pollen of plants. *Philosophical Magazine*, 4:161–173.
- [Burchfield et al., 2010] Burchfield, J. G., Lopez, J. A., Mele, K., Vallotton, P., and Hughes, W. E. (2010). Exocytotic vesicle behaviour assessed by total internal reflection fluorescence microscopy. *Traffic*, 11(4):429–439.
- [Burnham and Anderson, 2002] Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.
- [Byun et al., 2006] Byun, J., Verardo, M. R., Sumengen, B., Lewis, G. P., Manjunath, B. S., and Fisher, S. K. (2006). Automated tool for the detection of cell nuclei in digital microscopic images: Application to retinal images. *Molecular Vision*, 12:949–960.

- [Carrero et al., 2003] Carrero, G., McDonald, D., Crawford, E., de Vries, G., and Hendzel, M. J. (2003). Using FRAP and mathematical modeling to determine the in vivo kinetics of nuclear proteins. *Methods*, 29(1):14–28.
- [Carrington et al., 1995] Carrington, W. A., Lynch, R. M., Moore, E. D., Isenberg, G., Fogarty, K. E., and Fay, F. S. (1995). Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. *Science*, 268(5216):1483–1487.
- [Carslaw and Eger, 1959] Carslaw, H. S. and Eger, D. (1959). *Conduction of Heat in Solids*. Oxford University Press.
- [Cavanaugh et al., 2002] Cavanaugh, J. R., Bair, W., and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, 88:2530–2546.
- [Cha and Srihari, 2002] Cha, S.-H. and Srihari, S.-N. (2002). On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370.
- [Chalfie et al., 1994] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805.
- [Chandola et al., 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15.
- [Chen et al., 2006] Chen, Y., Lagerholm, C. B., Yang, B., and Jacobson, K. (2006). Methods to measure the lateral diffusion of membrane lipids and proteins. *Methods*, 39(2):147–153.
- [Chenouard et al., 2014] Chenouard, N., Smal, I., De Chaumont, E., Maska, M., Sbalzarini, I., Gon, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., R., C. A., J., G. W., Rohr, K., Kalaidzidis, Y., Liang, L., Duncan, J., Shen, H., Magnusson, K., Jalden, J., Paul-Gilloteaux, P., Roudot, P., Kervrann, C., Waharte, F., Tinevez, J.-Y., Willemse, J., Celler, K., Dan, H.-W., Tsai, Y.-S., Ortiz De Solorzano, C., Olivo-Marin, J.-C., and Meijering, E. (2014). Objective comparison of particle tracking methods. *Nature Methods*, 11:281–289.
- [Cheriyadat and Radke, 2008] Cheriyadat, A. M. and Radke, R. J. (2008). Detecting dominant motions in dense crowds. *J. Selected Topics in Sig. Processing*, 2(4):568–581.
- [Cherry, 1979] Cherry, R. J. (1979). Rotational and lateral diffusion of membrane proteins. *Biochimica et Biophysica Acta*, 559(4):289–327.

- [Chiani et al., 2003] Chiani, M., Dardari, D., and Simon, M. K. (2003). New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE Trans. Wireless Communications*, 2(4):840–845.
- [Chockalingam et al., 2013] Chockalingam, T., Emonet, R., and Odobez, J.-M. (2013). Localized anomaly detection via hierarchical integrated activity discovery. In *10th IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, Krakow.
- [Cinquin, 2011] Cinquin, B. (2011). *Étude dynamique en microscopie du rôle de Rab11a et de ses partenaires dans le recyclage des endosomes vers la membrane plasmique*. PhD thesis, Université Paris 7.
- [Clegg and Vaz, 1985] Clegg, R. M. and Vaz, W. L. C. (1985). Translational diffusion of proteins and lipids in artificial lipid bilayer membranes: A comparison of experiment with theory. *Progress in Protein-Lipid Interactions*, 1:173–229.
- [Cong et al., 2013] Cong, Y., Yuan, J., and Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition*, 46(7):1851–1864.
- [Cortes and Amit, 2008] Cortes, L. and Amit, Y. (2008). Efficient annotation of vesicle dynamics video microscopy. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(11):1998–2010.
- [Crivelli et al., 2011] Crivelli, T., Bouthemy, P., Cernuschi-Frías, B., and Yao, J.-F. (2011). Simultaneous motion detection and background reconstruction with a mixed-state conditional Markov random field. *Int. J. Computer Vision*, 94(3):295–316.
- [Crow, 1984] Crow, F. C. (1984). Summed-area tables for texture mapping. *ACM SIG-GRAPH Computer Graphics*, 18(3):207–212.
- [Cui et al., 2011] Cui, X., Liu, Q., Gao, M., and Metaxas, D. N. (2011). Abnormal detection using interaction energy potentials. In *2011 IEEE Conf. Computer Vision and Pattern Recognition*, pages 3161–3167.
- [Cupillard et al., 2002] Cupillard, F., Brémond, F., and Thonnat, M. (2002). *Tracking Group of People for Video Surveillance*. Video-Based Surveillance Systems. Kluwer Academic.
- [Deng et al., 2009] Deng, N., Xu, Y., Sun, D., Hua, P., Zheng, X., and Duan, H. (2009). Image processing for fusion identification between the GLUT4 storage vesicles and the plasma membrane. *J. Signal Processing Systems*, 54(1-3):115–125.
- [Deriche, 1990] Deriche, R. (1990). Fast algorithms for low-level vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(1):78–87.

- [Di Rienzo et al., 2014] Di Rienzo, C., Gratton, E., Beltram, F., and Cardarelli, F. (2014). Fast spatiotemporal correlation spectroscopy to determine protein lateral diffusion laws in live cell membranes. *Biophysical Journal*, 106(2):224a.
- [Einstein, 1905] Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560.
- [Einstein, 1956] Einstein, A. (1956). *Investigations on the Theory of the Brownian Movement*. Courier Corporation.
- [Embrechts et al., 1997] Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer-Verlag.
- [Escobar and Kornprobst, 2012] Escobar, M.-J. and Kornprobst, P. (2012). Action recognition via bio-inspired features: The richness of center-surround interaction. *Computer Vision and Image Understanding*, 116(5):593–605.
- [Feng et al., 2010] Feng, J., Zhang, C., and Hao, P. (2010). Online learning with self-organizing maps for anomaly detection in crowd scenes. In *20th Int. Conf. Pattern Recognition*, Istanbul.
- [Fick, 1855] Fick, A. (1855). Ueber diffusion. *Annalen der Physik*, 170(1):59–86.
- [Fortun et al., 2013] Fortun, D., Chen, C., Paul-Gilloteaux, P., Waharte, F., Salamero, J., and Kervrann, C. (2013). Correlation and variational approaches for motion and diffusion estimation in fluorescence imaging. In *2013 Proc. 21st European Signal Processing Conf.*, pages 1–5.
- [Fourier, 1822] Fourier, J. B. J. (1822). *Théorie Analytique de la Chaleur*. Chez Firmin Didot, père et fils.
- [François and Bouthemy, 1990] François, E. and Bouthemy, P. (1990). Derivation of qualitative information in motion analysis. *Image and Vision Computing Journal*, 8(4).
- [Gao et al., 2008] Gao, D., Mahadevan, V., and Vasconcelos, N. (2008). On the plausibility of the discriminant center-surround hypothesis for visual saliency. *J. of Vision*, 8(7):1–18.
- [Garate et al., 2014] Garate, C., Zaidenberg, S., Badie, J., and Bremond, F. (2014). Group tracking and behavior recognition in long video surveillance sequences. In *9th Int. Joint Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications*, Lisbon.

- [García-Sáez et al., 2010] García-Sáez, A. J., Carrer, D. C., and Schwille, P. (2010). Fluorescence correlation spectroscopy for the study of membrane dynamics and organization in giant unilamellar vesicles. In *Liposomes*, pages 493–508. Springer.
- [García-Sáez and Schwille, 2008] García-Sáez, A. J. and Schwille, P. (2008). Fluorescence correlation spectroscopy for the study of membrane dynamics and protein/lipid interactions. *Methods*, 46(2):116–122.
- [Gidon et al., 2012] Gidon, A., Bardin, S., Cinquin, B., Boulanger, J., Waharte, F., Héliot, L., de la Salle, H., Hanau, D., Kervrann, C., Goud, B., and Salamero, J. (2012). A Rab11A/myosin Vb/Rab11-FIP2 complex frames two late recycling steps of langerin from the ERC to the plasma membrane. *Traffic*, 13(6):815–833.
- [Guo et al., 2008] Guo, L., Har, J. Y., Sankaran, J., Hong, Y., Kannan, B., and Wohland, T. (2008). Molecular diffusion measurement in lipid bilayers over wide concentration ranges: A comparative study. *ChemPhysChem*, 9(5):721–728.
- [Hales et al., 2001] Hales, C. M., Griner, R., Hobdy-Henderson, K. C., Dorn, M. C., Hardy, D., Kumar, R., Navarre, J., Chan, E. K. L., Lapierre, L. A., and Goldenring, J. R. (2001). Identification and characterization of a family of Rab11-interacting proteins. *Journal of Biological Chemistry*, 276(42):39067–39075.
- [Hannig and Lee, 2006] Hannig, J. and Lee, T. C. M. (2006). On Poisson signal estimation under Kullback-Leibler discrepancy and squared risk. *J. Statistical Planning and Inference*, 136(3):882–908.
- [Hebert et al., 2005] Hebert, B., Costantino, S., and Wiseman, P. W. (2005). Spatiotemporal image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. *Biophysical Journal*, 88(5):3601–3614.
- [Hellriegel and Gratton, 2009] Hellriegel, C. and Gratton, E. (2009). Real-time multi-parameter spectroscopy and localization in three-dimensional single-particle tracking. *J. The Royal Society Interface*, 6(Suppl 1):S3–S14.
- [Hooke, 1665] Hooke, R. (1665). *Micrographia: Or Some Physiological Descriptions of Minutes Bodies made by Magnifying Glasses with Observations and Inquiries thereupon*. Royal Society.
- [Hoole, 1800] Hoole, S. (1800). *The Select Works of Antony Van Leeuwenhoek, Containing His Microscopical Discoveries in Many of the Works of Nature*, volume 1. G. Sidney.

- [Hozé et al., 2012] Hozé, N., Nair, D., Hosy, E., Sieben, C., Manley, S., Herrmann, A., Sibarita, J.-B., Choquet, D., and Holcman, D. (2012). Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging. *Proc. National Academy of Sciences*, 109(42):17052–17057.
- [Hu et al., 2008] Hu, M., Ali, S., and Shah, M. (2008). Learning motion patterns in crowded scenes using motion flow field. In *19th Int. Conf. Pattern Recognition*, Tampa.
- [Hutagalung and Novick, 2011] Hutagalung, A. H. and Novick, P. J. (2011). Role of Rab GTPases in membrane traffic and cell physiology. *Physiological Reviews*, 91(1):119–149.
- [Idrees et al., 2014] Idrees, H., Warner, N., and Shah, M. (2014). Tracking in dense crowds using prominence and neighborhood motion concurrence. *Image and Vision Computing*, 32(1):14–26.
- [Im et al., 2013] Im, K.-B., Schmidt, U., Kang, M.-S., Lee, J.-Y., Bestvater, F., and Wachsmuth, M. (2013). Diffusion and binding analyzed with combined point FRAP and FCS. *Cytometry Part A*, 83(9):876–889.
- [Jackson et al., 2011] Jackson, C., Glory, E., Murphy, R. F., and Kovačević, J. (2011). Model building and intelligent acquisition with application to protein subcellular location classification. *Bioinformatics*, 27(13):1854–1859.
- [Jacobson et al., 1987] Jacobson, K., Ishihara, A., and Inman, R. (1987). Lateral diffusion of proteins in membranes. *Annual Review of Physiology*, 49(1):163–175.
- [Jaqaman et al., 2008] Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, 5(5):695–702.
- [Jiang et al., 2007] Jiang, S., Zhou, X., Kirchhausen, T., and Wong, S. T. C. (2007). Detection of molecular particles in live cells via machine learning. *Cytometry Part A*, 71(8):563–575.
- [Jodoin et al., 2013] Jodoin, P.-M., Benezeth, Y., and Wang, Y. (2013). Meta-tracking for video scene understanding. In *2010 10th IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, pages 1–6.
- [Kapitza and Jacobson, 1986] Kapitza, H.-G. and Jacobson, K. A. (1986). Lateral motion of membrane proteins. In *Techniques for the Analysis of Membrane Proteins*, pages 345–375. Springer.

- [Kapur et al., 1985] Kapur, J. N., Sahoo, P. K., and Wong, A. K. C. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285.
- [Kenney and Keeping, 1947] Kenney, J. F. and Keeping, E. S. (1947). *Mathematics of Statistics*. Number Part I. Van Nostrand.
- [Kim et al., 2011] Kim, I.-H., Chen, Y.-C. M., Spector, D. L., Eils, R., and Rohr, K. (2011). Nonrigid registration of 2-D and 3-D dynamic cell nuclei images for improved classification of subcellular particle motion. *IEEE Trans. Image Processing*, 20(4):1011–1022.
- [Kim and Grauman, 2009] Kim, J. and Grauman, K. (2009). Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates. In *IEEE Conf. Computer Vision and Pattern Recognition*, Miami.
- [Kingman, 1992] Kingman, J. F. C. (1992). *Poisson Processes*, volume 3. Oxford university press.
- [Kolin and Wiseman, 2007] Kolin, D. L. and Wiseman, P. W. (2007). Advances in image correlation spectroscopy: Measuring number densities, aggregation states, and dynamics of fluorescently labeled macromolecules in cells. *Cell Biochemistry and Biophysics*, 49(3):141–164.
- [Kratz and Nishino, 2009] Kratz, L. and Nishino, K. (2009). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *IEEE Conf. Computer Vision and Pattern Recognition*, Miami.
- [Kratz and Nishino, 2012] Kratz, L. and Nishino, K. (2012). Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(5):987–1002.
- [Lapierre and Goldenring, 2005] Lapierre, L. A. and Goldenring, J. R. (2005). Interactions of Myosin Vb with Rab11 family members and cargoes traversing the plasma membrane recycling system. In *GTPases Regulating Membrane Targeting and Fusion*, volume 403 of *Methods in Enzymology*, pages 715–723. Academic Press.
- [Letinic et al., 2010] Letinic, K., Sebastian, R., Barthel, A., and Toomre, D. (2010). Deciphering subcellular processes in live imaging datasets via dynamic probabilistic networks. *Bioinformatics*, 26(16):2029–2036.
- [Li et al., 2013] Li, C., Han, Z., Ye, Q., and Jiao, J. (2013). Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. *Neurocomputing*, 119:94–100.

- [Li et al., 2015] Li, T., Chang, H., Wang, M., Ni, B., Hong, R., and Yan, S. (2015). Crowded scene analysis: A survey. *IEEE Trans. Circuits and Systems for Video Technology*, 25(3):367–386.
- [Li et al., 2014] Li, W., Mahadevan, V., and Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36(1):18–32.
- [Lindeberg, 1990] Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(3):234–254.
- [Lindeberg, 1993] Lindeberg, T. (1993). *Scale-space theory in computer vision*. Engineering and Computer Science: Robotics: Vision, Manipulation and Sensors. Springer.
- [Lindeberg, 1998] Lindeberg, T. (1998). Feature detection with automatic scale selection. *Int. J. Computer Vision*, 30(2):79–116.
- [Lindsay and McCaffrey, 2002] Lindsay, A. J. and McCaffrey, M. W. (2002). Rab11-FIP2 functions in Transferrin recycling and associates with endosomal membranes via its COOH-terminal domain. *Journal of Biological Chemistry*, 277(30):27193–27199.
- [Lu et al., 2013] Lu, C., Shi, J., and Jia, J. (2013). Abnormal event detection at 150 fps in matlab. In *2013 IEEE Int. Conf. Computer Vision*, pages 2720–2727.
- [Lucas and Kanade, 1981] Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc. Imaging Understanding Workshop*, volume 81, pages 121–130.
- [Macháň and Hof, 2010] Macháň, R. and Hof, M. (2010). Lipid diffusion in planar membranes investigated by fluorescence correlation spectroscopy. *Biochimica et Biophysica Acta*, 1798(7):1377–1391.
- [Magner, 2002] Magner, L. N. (2002). *A History of the Life Sciences, Revised and Expanded*. CRC Press.
- [Mallat, 1989] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(7):674–693.
- [Marian et al., 2007] Marian, A., Charrière, F., Colomb, T., Montfort, F., Kühn, J., Marquet, P., and Depeursinge, C. (2007). On the complex three-dimensional amplitude point spread function of lenses and microscope objectives: Theoretical aspects, simulations and measurements by digital holography. *J. Microscopy*, 225(2):156–169.

- [Maxwell, 1867] Maxwell, J. C. (1867). On the dynamical theory of gases. *Philosophical trans. Royal Society of London*, pages 49–88.
- [McDonald, 2014] McDonald, J. H. (2014). *Handbook of Biological Statistics*, volume 3. Sparky House Publishing.
- [McLachlan, 1992] McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Series in Probability and Statistics. Wiley.
- [Mehran et al., 2009] Mehran, R., Oyama, A., and Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *IEEE Conf. Computer Vision and Pattern Recognition*, Miami.
- [Meijering et al., 2012] Meijering, E., Dzyubachyk, O., and Smal, I. (2012). *Methods for cell and particle tracking*. Elsevier.
- [Mele et al., 2009] Mele, K., Coster, A., Burchfield, J. G., Lopez, J., James, D. E., Hughes, W. E., and Vallotton, P. (2009). Automatic identification of fusion events in TIRF microscopy image sequences. In *IEEE Int. Conf. Computer Vision Workshops*, pages 578–584.
- [Mendel, 1866] Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 44.
- [Merritt, 2013] Merritt, D. (2013). *Dynamics and Evolution of Galactic Nuclei*. Princeton University Press.
- [Michelman-Ribeiro et al., 2009] Michelman-Ribeiro, A., Mazza, D., Rosales, T., Stasevich, T. J., Boukari, H., Rishi, V., Vinson, C., Knutson, J. R., and McNally, J. G. (2009). Direct measurement of association and dissociation rates of DNA binding in live cells by fluorescence correlation spectroscopy. *Biophysical journal*, 97(1):337–346.
- [Miescher, 1871] Miescher, J. F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen. *Medicinish-chemische Untersuchungen*, 4:441–460.
- [Miklavc et al.,] Miklavc, P., Wittekindt, O. H., Felder, E., and Dietl, P. Ca²⁺-dependent Actin coating of lamellar bodies after exocytotic fusion: A prerequisite for content release or kiss-and-run, journal = *Annals of the New York Academy of Sciences*, volume = 1152, number = 1, publisher = Blackwell Publishing Inc, pages = 43–52, year = 2009.
- [Miura, 2005] Miura, K. (2005). *Tracking Movement in Cell Biology*, volume 95 of *Advances in Biochemical Engineering*. Springer Berlin Heidelberg.

- [Mo et al., 2014] Mo, X., Monga, V., Bala, R., and Fan, Z. (2014). Adaptive sparse representations for video anomaly detection. *IEEE Trans. Circuits and Systems for Video Technology*, 24(4):631–645.
- [Morgan et al., 1915] Morgan, T., Sturtevant, A. H., Muller, H. J., and Bridges, C. B. (1915). *The Mechanism of Mendelian Heredity*. Henry Holt.
- [Neil et al., 1997] Neil, M. A. A., Juškaitis, R., and Wilson, T. (1997). Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optical Letters*, 22(24):1905–1907.
- [Nguyen et al., 2015] Nguyen, H.-N., Kervrann, C., Cauchois, C., and Paveau, V. (2015). Automatic core segmentation and registration for fast tissue microarray de-arraying. In *2015 IEEE 12th Int. Symp. Biomedical Imaging*, pages 1439–1442, New York City.
- [Nielsen, 1999] Nielsen, H. B. (1999). Damping parameter in Marquardt’s method. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark.
- [Novick et al., 2006] Novick, P., Medkova, M., Dong, G., Hutagalung, A., Reinisch, K., and Grosshans, B. (2006). Interactions between Rab3, tethers, SNAREs and their regulators in exocytosis. *Biochemical Society Transactions*, 34(5):683–686.
- [Odobez and Bouthemy, 1995] Odobez, J.-M. and Bouthemy, P. (1995). Robust multiresolution estimation of parametric motion models. *J. Visual Communication and Image Representation*, 6:348–369.
- [Ohsugi et al., 2006] Ohsugi, Y., Saito, K., Tamura, M., and Kinjo, M. (2006). Lateral mobility of membrane-binding proteins in living cells measured by total internal reflection fluorescence correlation spectroscopy. *Biophysical J.*, 91(9):3456–3464.
- [Olivo-Marin, 2002] Olivo-Marin, J.-C. (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989–1996.
- [Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man, and Cybernetics*, 9(1):62–66.
- [Pearson, 1900] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50(302):157–175.
- [Pécot et al., 2015] Pécot, T., Bouthemy, P., Boulanger, J., Chessel, A., Bardin, S., Salamero, J., and Kervrann, C. (2015). Background fluorescence estimation and

- vesicle segmentation in live cell imaging with conditional random fields. *IEEE Trans. Image Processing*, 24(2):667–680.
- [Phansalkar et al., 2011] Phansalkar, N., More, S., Sabale, A., and Joshi, M. (2011). Adaptive local thresholding for detection of nuclei in diversity stained cytology images. In *2011 Int. Conf. Communications and Signal Processing*, pages 218–220.
- [Philibert, 2005] Philibert, J. (2005). One and a half century of diffusion: Fick, Einstein, before and beyond. *Diffusion Fundamentals*, 2(1):1–10.
- [Piciarelli et al., 2008] Piciarelli, C., Micheloni, C., and Foresti, G. L. (2008). Trajectory-based anomalous event detection. *IEEE Trans. Circuits and Systems for Video Technology*, 18(11):1544–1554.
- [Porikli and Haga, 2004] Porikli, F. and Haga, T. (2004). Event detection by eigenvector decomposition using object and frame features. In *IEEE Conf. Computer Vision and Pattern Recognition Workshop*, page 114.
- [Prydz et al., 2013] Prydz, K., Tveit, H., Vedeler, A., and Saraste, J. (2013). Arrivals and departures at the plasma membrane: Direct and indirect transport routes. *Cell and Tissue Research*, 352(1):5–20.
- [Ramadurai et al., 2009] Ramadurai, S., Holt, A., Krasnikov, V., van den Bogaart, G., Killian, A. J., and Poolman, B. (2009). Lateral diffusion of membrane proteins. *J. of the American Chemical Society*, 131(35):12650–12656.
- [Rehfeldt and Stichlmair, 2007] Rehfeldt, S. and Stichlmair, J. (2007). Measurement and calculation of multicomponent diffusion coefficients in liquids. *Fluid Phase Equilibria*, 256(1):99–104.
- [Reichert and Truskey, 1990] Reichert, W. M. and Truskey, G. A. (1990). Total internal reflection fluorescence (TIRF) microscopy. I. Modelling cell contact region fluorescence. *J. Cell Science*, 96(2):219–230.
- [Reid and Brunthaler, 2004] Reid, M. J. and Brunthaler, A. (2004). The proper motion of Sagittarius A*. II. The mass of Sagittarius A*. *The Astrophysical Journal*, 616(2):872–884.
- [Remak, 1852] Remak, R. (1852). Über extracellulare Entstehung thierischer Zellen und über die Vermehrung derselben durch Theilung. *Archiv für Anatomie, Physiologie und Wissenschaftliche Medicin*, pages 47–57.
- [Rezatofghi et al., 2012] Rezatofghi, S. H., Hartley, R., and Hughes, W. E. (2012). A new approach for spot detection in total internal reflection fluorescence microscopy. In *2012 IEEE Int. Symp. Biomedical Imaging*, Barcelona.

- [Rezatofghi et al., 2013] Rezatofghi, S. H., Pitkeathly, W. T. E., Gould, S., Hartley, R., Mele, K., Hughes, W. E., and Burchfield, J. G. (2013). A framework for generating realistic synthetic sequences of total internal reflection fluorescence microscopy images. In *Int. Symp. Biomedical Imaging*, pages 157–160.
- [Rizzoli and Jahn, 2007] Rizzoli, S. O. and Jahn, R. (2007). Kiss-and-run, collapse and ‘readily retrievable’ vesicles. *Traffic*, 8(9):1137–1144.
- [Rodriguez et al., 2009] Rodriguez, M., Ali, S., and Kanade, T. (2009). Tracking in unstructured crowded scenes. In *12th IEEE Int. Conf. Comp. Vis., ICCV’09, Kyoto*.
- [Rodriguez et al., 2011] Rodriguez, M., Sivic, J., Laptev, I., and J.-Y., A. (2011). Data-driven crowd analysis in videos. In *13th Int. Conf. Computer Vision, Barcelona*.
- [Roshtkhari and Levine, 2013] Roshtkhari, M. J. and Levine, M. D. (2013). An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Computer Vision and Image Understanding*, 117(10):1436–1452.
- [Ruusuvuori et al., 2010] Ruusuvuori, P., Äijö, T., Chowdhury, S., Garmendia-Torres, C., Selinummi, J., Birbaumer, M., Dudley, A. M., Pelkmans, L., and Yli-Harja, O. (2010). Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics*, 11(1).
- [Ruusuvuori et al., 2012] Ruusuvuori, P., Manninen, T., and Huttunen, H. (2012). Image segmentation using sparse logistic regression with spatial prior. In *20th European Signal Processing Conf.*, pages 2253–2257.
- [Ryan et al., 2011] Ryan, D., Denman, S., Fookes, C., and Sridharan, S. (2011). Textures of optical flow for real-time anomaly detection in crowds. In *8th IEEE Int. Conf. Advanced Video and Signal Based Surveillance, Klagenfurt*.
- [Sage et al., 2005] Sage, D., Neumann, F. R., Hediger, F., Gasser, S. M., and Unser, M. (2005). Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. *IEEE Trans. Image Processing*, 14(9):1372–1383.
- [Sahoo et al., 1997] Sahoo, P. K., Slaaf, D. W., and Albert, T. A. (1997). Threshold selection using a minimal histogram entropy difference. *Optical Engineering*, 36(7):1976–1981.
- [Sahoo et al., 1988] Sahoo, P. K., Soltani, S., and Wong, A. K. C. (1988). A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260.

- [Sako and Kusumi, 1994] Sako, Y. and Kusumi, A. (1994). Compartmentalized structure of the plasma membrane for receptor movements as revealed by a nanometer-level motion analysis. *J. Cell Biology*, 125(6):1251–1264.
- [Sarder and Nehorai, 2006] Sarder, P. and Nehorai, A. (2006). Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45.
- [Saxton and Jacobson, 1997] Saxton, M. J. and Jacobson, K. (1997). Single-particle tracking: Applications to membrane dynamics. *Annual Review of Biophysics and Biomolecular Structure*, 26(1):373–399.
- [Sbalzarini and Koumoutsakos, 2005] Sbalzarini, I. F. and Koumoutsakos, P. (2005). Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Structural Biology*, 151(2):182–195.
- [Scharf, 1990] Scharf, L. (1990). *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison Wesley.
- [Schleiden, 1838] Schleiden, M. J. (1838). Beiträge zur Phytogenesis. pages 137–176.
- [Schuss, 2012] Schuss, Z. (2012). The narrow escape problem – a short review of recent results. *J. Scientific Computing*, 53(1):194–210.
- [Schuss et al., 2007] Schuss, Z., Singer, A., and Holcman, D. (2007). The narrow escape problem for diffusion in cellular microdomains. *Proc. National Academy of Sciences*, 104(41):16098–16103.
- [Schwann, 1839] Schwann, T. (1839). *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen*. Sander.
- [Schwartz et al., 2007] Schwartz, S. L., Cao, C., Pylypenko, O., Rak, A., and Wandinger-Ness, A. (2007). Rab GTPases at a glance. *J. Cell Science*, 120(22):3905–3910.
- [Seiffert and Oppermann, 2005] Seiffert, S. and Oppermann, W. (2005). Systematic evaluation of FRAP experiments performed in a confocal laser scanning microscope. *J. Microscopy*, 220(1):20–30.
- [Sergeev, 2004] Sergeev, M. (2004). *High Order Autocorrelation Analysis in Image Correlation Spectroscopy*. PhD thesis, McGill University Montréal, Québec Canada.
- [Sezgin and Sankur, 2004] Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *J. Electronic Imaging*, 13(1):146–168.

- [Shafait et al., 2008] Shafait, F., Keysers, D., and Breuel, T. M. (2008). Efficient implementation of local adaptive thresholding techniques using integral images. In *15th Document Recognition and Retrieval Conference*, volume SPIE 6815, San Jose.
- [Sibarita, 2005] Sibarita, J.-B. (2005). Deconvolution Microscopy. In *Microscopy Techniques*, volume 95 of *Advances in Biochemical Engineering*, pages 201–243. Springer Berlin Heidelberg.
- [Singer et al., 2008] Singer, A., Schuss, Z., and Holcman, D. (2008). Narrow escape and leakage of Brownian particles. *Physical Review E*, 78(5):051111.
- [Sironi et al., 2011] Sironi, L., Solon, J., Conrad, C., Mayer, T. U., Brunner, D., and Ellenberg, J. (2011). Automatic quantification of microtubule dynamics enables RNAi-screening of new mitotic spindle regulators. *Cytoskeleton*, 68(5):266–278.
- [Smal et al., 2010] Smal, I., Loog, M., Niessen, W. J., and Meijering, E. H. W. (2010). Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Trans. Medical Imaging*, 29(2):282–301.
- [Smal et al., 2008] Smal, I., Niessen, W., and Meijering, E. (2008). A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In *5th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, pages 264–267.
- [Small and Stahlheber, 2014] Small, A. and Stahlheber, S. (2014). Fluorophore localization algorithms for super-resolution microscopy. *Nature methods*, 11(3):267–279.
- [Soille, 2003] Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer.
- [Solmaz et al., 2012] Solmaz, B., Moore, B. E., and Shah, M. (2012). Identifying behaviors in crowded scenes using stability analysis for dynamical systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(10):1–8.
- [Stauffer and Grimson, 2000] Stauffer, C. and Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):747–757.
- [Sternberg, 1983] Sternberg, S. R. (1983). Biomedical Image Processing. *IEEE Computer*, 16(1):22–34.
- [Steyer and Almers, 2001] Steyer, J. A. and Almers, W. (2001). A real-time view of life within 100 nm of the plasma membrane. *Nature Reviews Molecular Cell Biology*, 2:268–275.

- [Swaminathan et al., 1997] Swaminathan, R., Hoang, C. P., and Verkman, A. S. (1997). Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophysical Journal*, 72(4):1900–1907.
- [Tadin et al., 2006] Tadin, D., Lappin, J.-S., and Blake, R. (2006). Fine temporal properties of center-surround interactions in motion revealed by reverse correlation. *The Journal of Neuroscience*, 26(10):2614–2622.
- [ter Haar Romeny, 2003] ter Haar Romeny, B. M. (2003). *Front-end Vision and Multi-scale Image Analysis*. Computational Imaging and Vision. Springer.
- [Thida et al., 2013] Thida, M., Yong, Y. L., Climent-Pérez, P., Eng, H.-I., and Remagnino, P. (2013). A literature review on video analytics of crowded scenes. In *Intelligent Multimedia Surveillance*, pages 17–36. Springer.
- [Uzan-Gafsou et al., 2007] Uzan-Gafsou, S., Bausinger, H., Proamer, F., Monier, S., Lipsker, D., Cazenave, J.-P., Goud, B., de la Salle, H., Hanau, D., and Salamero, J. (2007). Rab11A controls the biogenesis of Birbeck granules by regulating Langerin recycling and stability. *Molecular Biology of the Cell*, 18(8):3169–3179.
- [Vaz et al., 1984] Vaz, W. L. C., Goodsaid-Zalduondo, F., and Jacobson, K. (1984). Lateral diffusion of lipids and proteins in bilayer membranes. *FEBS Letters*, 174(2):199–207.
- [Vincent, 1993] Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans. Image Processing*, 2(2):176–201.
- [Wang et al., 2011] Wang, X., Ma, K. T., Ng, G., and Grimson, W. E. L. (2011). Trajectory analysis and semantic region modeling using nonparametric hierarchical Bayesian models. *Int. J. Computer Vision*, 95(3):287–312.
- [Wiener, 1966] Wiener, N. (1966). Differential space. *J. Mathematical Physics*, 2:131–174.
- [Wilkinson and Schut, 1998] Wilkinson, M. H. F. and Schut, F. (1998). *Digital Image Analysis of Microbes: Imaging, Morphometry, Fluorometry and Motility Techniques and Applications*. John Wiley & Sons.
- [Wu et al., 2010] Wu, S., Moore, B. E., and Shah, M. (2010). Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In *IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco.

- [Wu et al., 2014] Wu, S., Wong, H.-S., and Yu, Z. (2014). A Bayesian model for crowd escape behavior detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(1):85–98.
- [Zhan et al., 2008] Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., and Xu, L.-Q. (2008). Crowd analysis: a survey. *Machine Vision and Applications*, 19:345–357.
- [Zhang et al., 2006] Zhang, B., Enninga, J., Olivo-Marin, J.-C., and Zimmer, C. (2006). Automated super-resolution detection of fluorescent rods in 2D. In *3rd IEEE Int. Symp. Biomedical Imaging: Nano to Macro*, pages 1296–1299.
- [Zhang et al., 2007] Zhang, B., Fadili, M. J., Starck, J.-L., and Olivo-Marin, J.-C. (2007). Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging. In *2007 14th IEEE Int. Conf. Image Processing*, San Antonio.
- [Zhao et al., 2011] Zhao, B., Fei-Fei, L., and Xing, E. P. (2011). Online detection of unusual events in videos via dynamic sparse coding. In *2011 IEEE Conf. Computer Vision and Pattern Recognition*, pages 3313–3320.
- [Zhou et al., 2012] Zhou, B., Tang, X., and Wang, X. (2012). Coherent filtering: Detecting coherent motions from crowd clutters. In *12th Eur. Conf. Computer Vision*, Firenze.
- [Zhou et al., 2013] Zhou, B., Tang, X., and Wang, X. (2013). Measuring crowd collectiveness. In *IEEE Conf. Computer Vision and Pattern Recognition*, Portland.
- [Zhou et al., 2011] Zhou, B., Wang, X., and Tang, X. (2011). Random field topic model for semantic region analysis in crowded scenes from tracklets. In *IEEE Conf. Computer Vision and Pattern Recognition*, Colorado Springs.
- [Zhu et al., 2014] Zhu, X., Liu, J., Wang, J., Li, C., and Lu, H. (2014). Sparse representation for robust abnormality detection in crowded scenes. *Pattern Recognition*, 47(5):1791–1799.

List of publications

- [Basset et al., 2014a] Basset, A., Boulanger, J., Bouthemy, P., Kervrann, C., and Salamero, J. (2014a). Sélection d'échelle automatique précise et seuillage localement adapté pour la segmentation de vésicules en microscopie TIRF. In *19ème congrès national sur la Reconnaissance de Formes et l'Intelligence Artificielle*, Rouen.
- [Basset et al., 2014b] Basset, A., Boulanger, J., Bouthemy, P., Kervrann, C., and Salamero, J. (2014b). SLT-LoG: A vesicle segmentation method with automatic scale selection and local thresholding applied to TIRF microscopy. In *2014 IEEE Int. Symp. Biomedical Imaging*, Beijing.
- [Basset et al., 2015a] Basset, A., Boulanger, J., Bouthemy, P., Kervrann, C., and Salamero, J. (2015a). Adaptive Spot Detection With Optimal Scale Selection in Fluorescence Microscopy Images. *IEEE Trans. Image Processing*, 24(11):4512–4527.
- [Basset et al., 2015b] Basset, A., Boulanger, J., Bouthemy, P., Kervrann, C., and Salamero, J. (2015b). Détection de spots avec sélection d'échelle automatique et seuillage adaptatif en microscopie de fluorescence. *Traitement du Signal*, 32(2–3).
- [Basset et al., 2014c] Basset, A., Bouthemy, P., Boulanger, J., Salamero, J., and Kervrann, C. (2014c). Detection and classification of dynamic subcellular events in TIRF microscopy sequences. In *2014 IEEE Int. Symp. Biomedical Imaging*, Beijing.
- [Basset et al., 2015c] Basset, A., Bouthemy, P., Boulanger, J., Waharte, F., Kervrann, C., and Salamero, J. (2015c). Detection and estimation of membrane diffusion events during exocytosis in TIRFM image sequences. In *IEEE Int. Symp. Biomedical Imaging*, New York City.
- [Basset et al., 2013a] Basset, A., Bouthemy, P., and Kervrann, C. (2013a). Classification instantanée de mouvements de foules dans des vidéos. In *XXIVe Colloque Gretsi*, Brest.
- [Basset et al., 2013b] Basset, A., Bouthemy, P., and Kervrann, C. (2013b). Frame-by-frame crowd motion classification from affine motion models. In *10th IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, pages 282–287, Kraków.

- [Basset et al., 2014d] Basset, A., Bouthemy, P., and Kervrann, C. (2014d). Recovery of motion patterns and dominant paths in videos of crowded scenes. In *IEEE Int. Conf. Image Processing*, Paris.
- [Pérez-Rúa et al., 2015] Pérez-Rúa, J.-M., Basset, A., and Bouthemy, P. (2015). Local abnormal event detection in crowd videos from histograms of labeled affine flows. *Submitted to IEEE Trans. Pattern Analysis and Machine Intelligence*.

Acronyms

- AIC** Akaike information criterion. 121, 129
- AICc** Akaike information criterion corrected for small sample size. 137, 142
- ATLAS** adaptive thresholding of LoG images with auto-selected scale. 28, 29, 41, 88
- BCCE** brightness constancy constraint equation. 103
- C-CRAFT** conditional random fields for protein transport carriers segmentation. 46, 47, 55, 58
- CDF** cumulative density function. 73
- CFAR** constant false alarm rate. 134
- CTM** correlated topic models. 118
- DFD** displaced frame difference. 90, 119
- DFT** discrete Fourier transform. 35
- DNA** deoxyribonucleic acid. 12–15
- DOF** depth of field. 28
- FCS** fluorescence correlation spectroscopy. 99, 101, 104
- FPR** false positive rate. 92, 137, 138, 141, 142
- FRAP** fluorescence recovery after photobleaching. 70, 76, 77, 102, 104, 106
- PROC** free receiver-operator characteristic. 47, 52, 54, 55, 58
- GAT** generalized Anscombe transform. 23, 46
- GEV** generalized extreme value. 89, 90, 92, 137
- GFP** green fluorescent protein. 13, 27
- HD** *h*-dome. 30, 46, 47, 58
- IRLS** iteratively reweighted least squares. 119
- IUWT** iteratively undecimated wavelet transform. 30
- LAF** labeled affine flow. 116, 117, 122, 142, 144
- LCSS** longest common subsequence. 118
- LoG** Laplacian of Gaussian. 28–30, 35–37, 43, 50, 54, 59, 60, 62, 95
- LR-MRF** logistic regression with Markov random field. 50

- MALE** mean absolute logarithmic error. 106
- ML** maximum likelihood. 133
- MPG** mixed Poisson-Gaussian. 22, 23
- MPHD** maximum possible height dome. 31, 46, 52, 54, 55, 59
- MSD** mean square displacement. 100
- MS-VST** multiscale variance stabilizing transform. 23, 30, 46, 47, 52, 55, 58–60
- MTH** morphological top hat filter. 30
- PDE** partial differential equation. 74
- PDF** probability density function. 72, 74
- PFA** probability of false alarm. 28, 32, 44, 47, 62, 89, 95, 137, 141, 142
- PSF** point spread function. 19, 20, 71, 74, 75, 77, 80, 81, 83, 85, 108
- PSNR** peak signal-to-noise ratio. 52, 54, 58, 91, 92
- RNA** ribonucleic acid. 14, 15
- ROC** receiver-operator characteristic. 137, 138, 140, 141
- SEF** spot enhancing filter. 29
- SIM** structured illumination microscopy. 13
- SNR** signal-to-noise ratio. 18, 19, 29, 30, 38, 41, 48, 52, 106, 107
- SPT** single particle tracking. 19, 99, 100
- SSED** “small-extent source with exponential decay release”. 73, 82, 83, 86, 97, 101, 102, 104, 107, 108
- STICS** spatiotemporal image correlation spectroscopy. 103
- TfR** Transferrin receptor. 4, 6, 8, 11, 17, 28, 58, 65–67, 69, 82, 83, 87, 94, 99, 108, 111, 143
- TH** top hat filter. 29, 30
- TICS** temporal image correlation spectroscopy. 100
- TIRFM** total internal reflection fluorescence microscopy. 1, 5, 8, 11, 17–20, 28, 29, 33–35, 38, 46, 50, 52, 55, 59, 62, 65, 72, 77, 81, 82, 86, 88, 91, 92, 95, 102, 104, 106, 108, 111, 143
- TMA** tissue micro-array. 59
- TPR** true positive rate. 92, 137, 138
- WMP** wavelet multiscale product. 30