DISSERTATION

submitted

to the

Combined Faculties for the Natural Sciences and for Mathematics

of the

Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

William J. Godinez, M.Sc.

Born in: Guatemala City, Guatemala

Oral examination: .....

Probabilistic Tracking and Behavior Identification of Fluorescent Particles

Advisor: PD Dr. Karl Rohr

Abstract

Explicit and tractable characterizations of the dynamical behavior of virus particles are pivotal for a thorough understanding of the infection mechanisms of viruses. This thesis deals with the problem of extracting symbolic representations of the dynamical behavior of fluorescent particles from fluorescence microscopy image sequences. The focus is on the behavior of virus particles such as fusion with the cell membrane. A numerical representation is obtained by tracking the particles in the image sequences. We have investigated probabilistic tracking approaches, including approaches based on the Kalman filter as well as based on particle filters. For reasons of efficiency and robustness, we developed a tracking approach based on the probabilistic data association (PDA) algorithm in combination with an ellipsoidal sampling scheme that exploits effectively the image data via parametric appearance models. To track objects in close proximity, we compute the support that each image position provides to each tracked object relative to the support provided to the object's neighbors. After tracking, the problem of mapping the trajectory information computed by the tracking approaches to symbolic representations of the behavior arises. To compute symbolic representations of behaviors related to the fusion of single virus particles with the cell membrane based on their intensity over time, we developed a layered probabilistic approach based on stochastic hybrid systems as well as hidden Markov models (HMMs). We use a maxbelief strategy to efficiently combine both representations. The layered approach describes the intensity, intensity models, and behaviors of single virus particles. We introduce models for the evolution of the intensity and the behavior. To compute estimates for the intensity, intensity models, and behaviors we use a hybrid particle filter and the Viterbi algorithm. The de-

veloped approaches have been applied to synthetic images as well as to real microscopy image sequences displaying human immunodeficiency virus (HIV-1) particles. We have performed an extensive quantitative evaluation of the performance and a comparison with several existing approaches. It turned out that our approaches outperform previous ones, thus yielding more accurate and more reliable information about the behavior of virus particles. Moreover, we have successfully applied our tracking approaches to 3D image sequences displaying herpes simplex virus (HSV) replication compartments. We also applied the tracking approaches to image data displaying microtubule tips and analyzed their motion. In addition, our tracking approaches were successfully applied to the 2D and 3D image data of a Particle Tracking Challenge.

Zusammenfassung

Explizite und traktable Charakterisierungen des dynamischen Verhaltens von Viruspartikeln sind entscheidend für ein gründliches Verständnis der Infektionsmechanismen von Viren. Diese Dissertation beschäftigt sich mit der Extraktion von symbolischen Repräsentationen des dynamischen Verhaltens von fluoreszierenden Partikeln aus Fluoreszenz-Mikroskopie Bildsequenzen. Der Schwerpunkt liegt auf dem Verhalten von Viruspartikeln, wie beispielweise der Fusion einzelner Viruspartikel mit der Zellmembran. Eine numerische Repräsentation wird durch Tracking (Verfolgung) von Partikeln erhalten. Wir haben probabilistische Trackingverfahren einschließlich Verfahren basierend auf dem Kalman-Filter und basierend auf dem Partikel-Filter untersucht. Aus Gründen der Recheffizienz und Robustheit haben wir ein Trackingverfahren entwickelt basierend auf dem Probabilistic Data Association (PDA) Algorithmus in Kombination mit einem ellipsoidischen Abtastschema, das die Bilddaten durch parametrische Intensitätsmodelle effektiv auswertet. Um Objekte zu verfolgen, die nah beieinander sind, berechnen wir die Unterstützung, die jede Bildposition zu jedem verfolgten Objekt gibt, relativ zu der Unterstützung, die zu den Nachbarn des Objektes gegeben wird. Nach dem Tracking wird die Trajektorieninformation einer symbolischen Repräsentation des Verhaltens zugeordnet. Um symbolische Repräsentationen des Verhaltens der Fusion von einzelnen Viruspartikeln mit der Zellmembran basierend auf der Bildintensität über der Zeit zu berechnen, haben wir ein geschichtetes probabilistisches Verfahren basierend auf stochastischen hybriden Systemen und Hidden Markov Modellen (HMM) entwickelt. Wir verwenden eine maxbelief Strategie, um beide Repräsentationen zu kombinieren. Das

geschichtete Verfahren beschreibt die Intensität, Intensitäts-Modelle, und das Verhalten einzelner Viruspartikel. Wir stellen Modelle für die Entwicklung der Intensität und des Verhaltens vor. Um Schätzungen der Variablen zu bestimmen, benutzen wir einen hybriden Partikel-Filter und den Viterbi Algorithmus. Die entwickelten Verfahren wurden sowohl für synthetische Bilddaten als auch für reale Fluoreszenz-Mikroskopie Bilddaten, die Humane Immundefizienz-Virus (HIV-1) Partikel zeigen, angewendet. Wir haben eine umfangreiche quantitative Evaluierung der Performanz durchgeführt sowie einen Vergleich mit mehreren existierenden Verfahren. Es hat sich gezeigt, dass unsere Verfahren bessere Ergebnisse liefern als bisherige Verfahren, daher ergeben sich genauere und zuverlässigere Information über das Verhalten von Viruspartikeln. Ferner haben wir unsere Trackingverfahren erfolgreich auf 3D Bildsequenzen angewendet, die Replikationskompartimente von Herpes-Simplex-Viren (HSV) zeigen. Wir haben die Trackingverfahren auch auf Bildsequenzen angewendet, die Mikrotubuli-Enden zeigen und haben deren Bewegung analysiert. Außerdem wurden die Trackingverfahren erfolgreich auf die 2D und 3D Bilddaten einer Particle Tracking Challenge angewendet.

Acknowledgments

I thank God.

I want to thank Karl Rohr for his wise guidance and support throughout all these years since my university studies. He introduced me to the field of computer vision, and for that I am vastly thankful. My sincere gratitude also goes to Christoph Schnörr for agreeing to be a member of the examination committee.

I want to express my gratitude to all the people with whom I had very productive collaborations. I want to thank Marko Lampe, Peter Koch, Abdul Rahman, and Barbara Müller from the Virology Department at the University of Heidelberg for a fruitful and fun collaboration. The experimental results presented in this thesis are largely based on Marko's and Peter's image data. I also thank Maik J. Lehmann, also from the Virology Department, for answering my questions about the different fluorescent labeling strategies. The collaboration with Sergey Ivanchenko, Viola Baumgärtel, and Don C. Lamb from LMU Munich was also a great experience and I thank them for that. I also wish to thank Lynne Chang and David M. Knipe at Harvard Medical School for the successful collaboration. I am also thankful to Astrid Marx, Pete Bankhead, and Ulrike Engel from the Nikon Imaging Center for our joint effort on microtubule tracking. I also thank Konstantinos Lymperopoulos and Dirk-Peter Herten from the Institute for Physical Chemistry at the University of Heidelberg for their cutting-edge single-molecule image data. Many thanks to Barbara Di Ventura from the University of Heidelberg for the enthusiastic joint work on chromosome tracking. I also want to thank Fabian Rengier and Stefan Wörz from the University of Heidelberg, with whom I worked on medical images. This was also very interesting and enjoyable.

I also want to thank my colleagues from the Biomedical Computer Vision (BMCV) group. I am very grateful to Nathalie Harder and Stefan Wörz with whom I shared the office for many years. I am thankful for the great office environment and also for the great camaraderie. I am thankful to Il-Han Kim, Petr Matula, and Andreas Biesdorf for our enjoyable technical discussions. I also want to thank Astha Jaiswal and Vasil Tsimashchuk for our joint work on tracking. I thank Marco Tektonidis for his great 3D registration results. I also shared the office with Wei Liao and Jan-Philip Bergeest, and

I am grateful for the insightful and fun conversations. I also must thank Simon Eck for the valuable cycling tips. Jorge Jara from Universidad de Chile visited our group for a few months and I thank him for his friendship and conversations in Spanish.

I also want to thank all the people from the Theoretical Bioinformatics division at the DKFZ Heidelberg, in particular, Roland Eils. I also want to thank Manuela Schäfer for the administrative support and Peter Weyrich for the IT support. I also thank the folks at the BIOQUANT with whom I shared an enjoyable research environment.

My final thanks are reserved for my Dad, my Mom, and my Brother. Many thanks for your love, encouragement, and steadfast support.

The work described in this thesis was carried out as part of my PhD studies in Computer Science at the Faculty of Mathematics and Computer Science of the University of Heidelberg. The work was supported by the BMBF (FORSYS) project VIROQUANT and the EU project SysPatho (FP7).

Contents

List of Figures

List of Tables

Publications

Journal Papers

- A. Marx, **W.J. Godinez**, V. Tsimashchuk, P. Bankhead, K. Rohr, and U. Engel, "*Xenopus* cytoplasmic linker associated protein 1 (XCLASP1) promotes axon elongation and advance of pioneer microtubules," *Molecular Biology of the Cell*, 2013, in press.

- **W.J. Godinez**, M. Lampe, P. Koch, R. Eils, B. Müller, and K. Rohr, "Identifying virus-cell fusion in two-channel fluorescence microscopy image sequences based on a layered probabilistic approach," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1786–1808, 2012.

- L. Chang, **W.J. Godinez**, I.-H. Kim, M. Tektonidis, P. de Lanerolle, R. Eils, K. Rohr, and D. M. Knipe, "Herpesviral replication compartments move and coalesce at nuclear speckles to enhance export of viral late mRNA," *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. E136–E144, 2011.

- F. Rengier, S. Wörz, **W.J. Godinez**, H. Schumacher, D. Böckler, K. Rohr, H.-U. Kauczor, and H. von Tengg-Kobligk, "Development of in-vivo quantitative geometric mapping of the aortic arch for advanced endovascular aortic repair: feasibility and preliminary results," *Journal of Vascular and Interventional Radiology*, vol. 22, no. 7, pp. 980–986, 2011.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences," *Medical Image Analysis*, vol. 13, pp. 325–342, 2009.

- S. Ivanchenko, **W.J. Godinez**, M. Lampe, H.-G. Kräusslich, R. Eils, K. Rohr, C. Bräuchle, B. Müller, and D. C. Lamb, "Dynamics of HIV-1 assembly and release," *PLoS Pathogens*, vol. 5, no. 11, p. e1000652, 2009.

- P. Koch, M. Lampe, **W.J. Godinez**, B. Müller, K. Rohr, H.-G. Kräusslich, and M. J. Lehmann, "Visualizing fusion of pseudotyped HIV-1 particles in real time by live cell microscopy," *Retrovirology*, vol. 6, no. 1, p. 84, 2009.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, A. Wünsche, R. Eils, J. Ellenberg, and K. Rohr, "Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time," *Genome Research*, vol. 19, no. 11, pp. 2113–2124, 2009.

Conference Papers

- V. Tsimashchuk, **W.J. Godinez**, M. J. Lehmann, and K. Rohr, "Analysis of viral surfing based on fluorescence microscopy imaging and automatic tracking," in *Proc. 9th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, pp. 378–381, May 2012.

- A. Jaiswal, **W.J. Godinez**, R. Eils, M. J. Lehmann, and K. Rohr, "Tracking virus particles in fluorescence microscopy images using two-step multi-frame association," in *Proc. 9th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, pp. 664–667, May 2012.

- A. Jaiswal, **W.J. Godinez**, R. Eils, M. J. Lehmann, and K. Rohr, "Tracking virus particles in microscopy images using multi-frame association," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'12)*, Berlin, Germany, pp. 231–236, March 2012.

- **W.J. Godinez**, M. Lampe, R. Eils, B. Müller, and K. Rohr, "Tracking multiple particles in fluorescence microscopy images via probabilistic data association," in *Proc. 8th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'11)*, Chicago, IL, USA, pp. 1925–1928, March–April 2011.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, A. Wünsche, J. Ellenberg, R. Eils, and K. Rohr, "Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'10)*, Aachen, Germany, pp. 81–85, March 2010.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Performance evaluation of approaches for tracking multiple virus particles in fluorescence microscopy image sequences," in *Proc. 4th Internat. Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB'09)*, Bethesda, MD, USA, September 2009.

- **W.J. Godinez**, M. Lampe, S. Wörz, R. Eils, B. Müller, and K. Rohr, "Identifying fusion events in fluorescence microscopy images," in *Proc. 6th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston, MA, USA, pp. 1170–1173, June–July 2009.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Evaluation of approaches for tracking virus particles in fluorescence microscopy images," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'09)*, Heidelberg, Germany, pp. 207–211, March 2009.

- S. Wörz, **W.J. Godinez**, and K. Rohr, "Segmentation of 3D tubular structures based on 3D intensity models and particle filter tracking," in *Proc. SPIE Medical Imaging 2009: Image Processing*, Lake Buena Vista, FL, USA, pp. 72591P–72591P–8, February 2009.

- S. Wörz, **W.J. Godinez**, and K. Rohr, "Probabilistic tracking and model-based segmentation of 3D tubular structures," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'09)*, Heidelberg, Germany, pp. 41–45, March 2009.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Probabilistic tracking of virus particles in fluorescence microscopy images," in *Proc. 5th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, pp. 272–275, May 2008.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Probabilistic tracking of virus particles in fluorescence microscopy image sequences," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'08)*, Berlin, Germany, pp. 448–452, April 2008.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Tracking of virus particles in time-lapse fluorescence microscopy image sequences," in *Proc. 4th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07)*, Arlington, VA, USA, pp. 256–259, April 2007.

- **W.J. Godinez**, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Tracking of virus particles in time-lapse fluorescence microscopy image sequences," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'07)*, Munich, Germany, pp. 6–10, March 2007.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, J. Ellenberg, R. Eils, and K. Rohr, "Determination of mitotic delays in 3D fluorescence microscopy images of human

cells using an error-correcting finite state machine," in *Proc. 4th IEEE Internat. Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07)*, Arlington, VA, USA, pp. 1044–1047, April 2007.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, J. Ellenberg, R. Eils, and K. Rohr, "Determination of mitotic delays in 3D fluorescence microscopy images of human cells using an error-correcting finite state machine," in *Proc. Workshop Bildverarbeitung für die Medizin (BVM'07)*, Munich, Germany, pp. 242–246, March 2007.

- I.-H. Kim, **W.J. Godinez**, N. Harder, F. Mora-Bermúdez, J. Ellenberg, R. Eils, and K. Rohr, "Compensation of global movement for improved tracking of cells in time-lapse confocal microscopy image sequences," in *Proc. SPIE Medical Imaging 2007: Image Processing*, San Diego, CA, USA, p. 65121R, February 2007.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, J. Ellenberg, R. Eils, and K. Rohr, "Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences," in *Proc. Ninth Internat. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)*, Copenhagen, Denmark, vol. 4190, pp. 840–848, October 2006.

Book Chapters

- K. Rohr, **W.J. Godinez**, N. Harder, S. Wörz, J. Mattes, W. Tvaruskó, and R. Eils, Tracking and Quantitative Analysis of Dynamic Movements of Cells and Particles, in R. D. Goldman, J. R. Swedlow, and D. L. Spector (Eds.), *Live cell imaging - a laboratory manual*, pp. 239–256. Woodbury, New York, USA: Cold Spring Harbor Laboratory Press, 2nd ed., 2010.

- N. Harder, F. Mora-Bermúdez, **W.J. Godinez**, J. Ellenberg, R. Eils, and K. Rohr, Automated analysis of the mitotic phases of human cells in 3-D fluorescence microscopy image sequences, in J. Rittscher, R. Machiraju, and S. T. C. Wong (Eds.), *Microscopic image analysis for life science applications*, pp. 283–294. Norwood, Massachusetts, USA: Artech House, 2008.

Chapter 1

Introduction

1.1 Motivation

Infections caused by viruses generally represent a serious threat to human life. The prejudicial potential of viral infections is highlighted by pandemic outbreaks of viruses such as the H1N1 influenza A virus (e.g., [180]), the human-immunodeficiency virus type 1 (HIV-1) (e.g., [188]), as well as the hepatitis C virus (e.g., [112]). Entry of viruses into cells is crucial for infection. In general, an enveloped *virus particle*, such as HIV-1, can enter a cell via two alternative mechanisms: endocytosis or fusion at the plasma membrane ([121], [197]). In the former case, the cell internalizes the virus particle by engulfing it into an intracellular vesicle. In the latter case, the virus particle directly fuses with the plasma membrane. However, the dynamics of these entry processes are not well understood. Detailed insight into the fusion process results in a better understanding of the infectivity of the viruses, which in turn can lead to the development of more effective antiviral drugs. In this thesis we focus on the fusion mechanisms of pseudotyped human immunodeficiency virus type 1 (HIV-1) particles with the cell membrane using live-cell microscopy as well as double fluorescent labeling strategies ([128], [108], [103]). Concretely, each virus particle is tagged with two different fluorescent labels: one label is attached to the outer shell of the particle while the other is attached to its inner part. This labeling technique results in two-channel fluorescence microscopy image sequences. Upon fusion with the cell membrane, the particle's outer shell is disrupted and the inner part is released into the cell. Fusion leads to the disappearance of the fluorescent label attached to the outer shell because this label is diluted into the cell membrane, and the label attached to the inner part remains visible. Hence, fusion can be defined as a transient behavior that is described by changes in the fluorescence intensities of a single particle. Time-lapse fluorescence microscopy allows a detailed observation of such changes. However, the resulting fluorescence image sequences provide

an *implicit* representation of the behavior of the virus particles. To obtain a tractable description of the behavior, *explicit* representations of the behavior are required, and to draw statistically sound conclusions about the behavior, a large number of virus particles must be observed. This leads to a large amount of image data from which explicit representations of the behavior are to be extracted. Manual extraction of such explicit representations is tedious and error-prone. In contrast, automatic extraction via *computer vision* approaches provides efficiency and reproducibility. This thesis presents a solution grounded within a Bayesian framework for automatically estimating behaviors of virus particles from fluorescence microscopy image sequences. While the focus is on virus particles, we also demonstrate the applicability of the developed approaches to other types of fluorescent structures (e.g., microtubule tips).

1.2 Estimating Behaviors of Fluorescent Particles

In this section, we consider the problems of estimating behaviors of fluorescent particles in general by briefly discussing the challenging properties of the image data acquired via time-lapse fluorescence microscopy. Next, we elaborate on the image analysis tasks involved in estimating behaviors of fluorescent particles using time-lapse microscopy image data.

1.2.1 Time-Lapse Fluorescence Microscopy

In time-lapse fluorescence microscopy, the subcellular structures of interest are labeled with a fluorescent protein (e.g., the green fluorescent protein, GFP) that emits light after excitation with a light source (e.g., a laser) [117]. Obtaining images of the structures in a temporally sequential manner allows the observation of the structures over time. While fluorescence microscopy is a versatile technique for monitoring the behavior of subcellular structures, the technique presents certain challenges. One problem relates to both the optical limitations of typical fluorescent microscopes and the small size of the studied structures. For example, storage vesicles containing glucose transporter 4 (GLUT4) located in muscle cells have a size in the range of 50-100 nanometers [81]. Human-immunodeficiency virus type 1 (HIV-1) particles have a size of ca. 140 nm. Since the optical system of typical microscopes has a limited spatial resolution (ca. 200 nm laterally, i.e., in the focal plane [80]), small structures cannot be spatially resolved. Instead, these point-like structures appear distorted, where the distortion is described as a convolution between the object and the point spread function (PSF) of the optical system of the microscope [2]. In several fluorescence microscopes (e.g., wide field fluorescence microscope, laser scanning confocal microscope), the PSF can

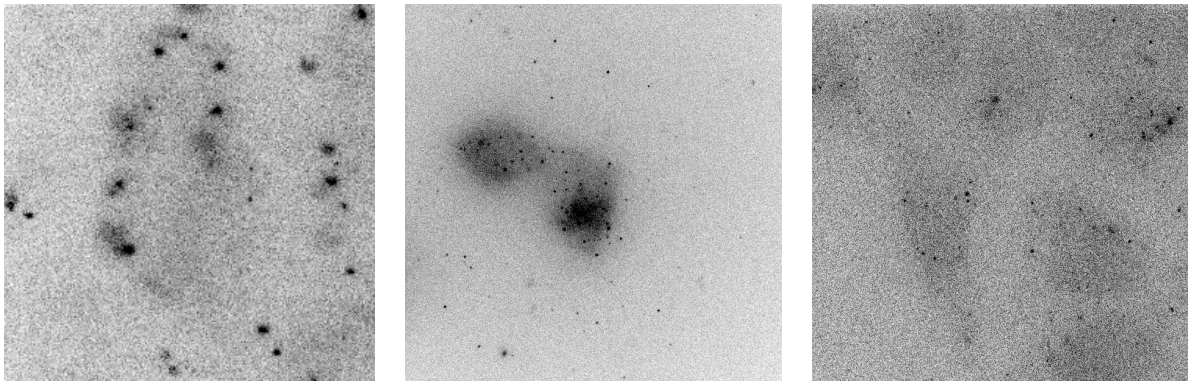


Figure 1.1: Sample images from three real microscopy image sequences displaying HIV-1 particles (black spots). It can be seen that the images differ quite a lot. (Image intensities have been inverted for visualization purposes).

be well approximated by a Gaussian function [222]. A sub-resolution object appears in the image as a Gaussian-like *fluorescent particle*. As an example, Figure 1.1 presents three images displaying HIV-1 particles. The particles appear as black spots because the image intensities have been inverted. The Gaussian-like appearance of the spots is parametrized by the peak intensity and the standard deviation of the underlying Gaussian function. Besides these two parameters, the Gaussian-like appearance reveals no additional visual cues (e.g., texture, true shape) about the underlying structures. Often the structures look very similar and thus identifying each particular structure over time in an image sequence is difficult. Another consequence of the small size of the structures is that a limited number of fluorescent molecules can be attached to them. This leads to a relatively low fluorescence intensity. Since some structures within the cell (e.g., mitochondria, lysosomes) are naturally fluorescent [7], [16], the cellular background also emits some light. The low fluorescence intensity of the studied structures in conjunction with the *cellular autofluorescence* leads to a low image contrast for the particles. Another problem in time-lapse fluorescence microscopy is the process whereby the ability of fluorescent proteins to emit light is damaged by the light cast onto them (i.e., the incident light) [182]. This process, known as *photobleaching*, is of particular relevance when acquiring a sequence of images in a temporally sequential manner, where light is shed upon the fluorescent structures every time an image is acquired. Over time, photobleaching leads to a decrease in the fluorescence intensity of the structure. Photobleaching has also implications on the rate at which the images are acquired. Higher acquisition rates imply acquiring a large number of images over a relatively short period of time. Because of photobleaching, the time during which the fluorescent proteins remain visible is rather short and this prevents observing the behavior of the structures

over a long period of time. Conversely, low acquisition rates may enable the observation of the structures over longer periods. However the behavior of the structures between the acquisition of two consecutive images is not observed. That is, such a low temporal resolution may not provide a rich enough description of the dynamical behavior of the structures. A compromise between the two acquisition strategies (i.e., high and low acquisition rates) may be reached by reducing the intensity of the incident light, which reduces the degradation of the fluorescent proteins. A low intensity light source is also needed to prevent undesirable alterations of certain cellular processes, such as the cell's circadian rhythm. One drawback of using a low intensity light source is that the emitted fluorescence intensity of the structures is quite low. A low level of fluorescence intensity is given by a low number of emitted photons. The number of emitted photons over a certain period of time is a random process modeled using a Poisson distribution, where the mean and standard deviation of the distribution are given by the average number of photons and the square root of the average number of photons, respectively [187], [192]. Taking the average number of photons as the 'signal' and the square root of the average number of photons as the 'noise', the *signal-to-noise ratio* (SNR) is essentially the square root of the average number of photons. Thus, the low level of fluorescence intensity induced by the low intensity light source that is required to prevent the degradation of the fluorescent proteins leads to a low SNR. A low SNR is evident in the images shown in Figure 1.1. Certainly there are other sources of noise in fluorescence microscopy, such as leakage current caused by a poor insulation of the photomultiplier tube that is used for photon counting in certain microscope setups. These distortions are generally related to the shortcomings of the items involved in the microscopy system. To summarize, the study of subcellular structures via time-lapse fluorescence microscopy is circumscribed by certain challenges, such as the optical resolution of the microscope system, photobleaching, as well as the inherent image noise.

1.2.2 Image Analysis Tasks and Challenges

Given a microscopy image sequence displaying multiple fluorescent particles, the task of estimating dynamical behaviors of subcellular structures entails transforming the given sequence into labels that provide a concrete interpretation of the behavior. Concretely, images within a time-lapse microscopy image sequence are comprised by pixel data associated with intensity values that provide only an implicit indication of the behavior of the individual structures. The aim is to transform the pixel data into an explicit description of the behavior. This task can be approached by deriving descriptions of the behavior of the individual structures at different levels of abstraction in a bottom-up manner. Concretely, first the structures of interest are *localized* within the images. Sec-

ond, using the calculated locations of the structures, the position over time of individual structures is determined, i.e., the individual structures are *tracked*. Once the position over time of the structures is calculated, the temporal statistics (e.g., displacement, intensity over time, size over time) of the individual structures are computed. Based on these statistics, explicit descriptions of the dominant *behaviors* underlying the temporal statistics are obtained. Thus, within such an approach, the problem of estimating dynamical behaviors of subcellular structures based on fluorescence microscopy image sequences can be decomposed into three tasks: (1) object localization, (2) tracking, as well as (3) behavior identification. In general, these computer vision tasks are not trivial. In fluorescence microscopy, the tasks must take into account the particular limitations of this imaging technique, too. We discuss briefly these tasks and the challenges entailed by time-lapse fluorescence microscopy images.

The task of *object localization* usually involves two steps: detection as well as localization. The goal of detection is to determine image regions that correspond to the objects of interest (in our case fluorescent particles). The detection of fluorescent particles is mainly limited by the small size of the particles, which typically exhibit a Gaussian-like appearance with a low intensity contrast. The cellular autofluorescence is also problematic for object detection, since it reduces the contrast. Likewise, spurious fluorescent compounds may lead to particle-like artefacts. The image noise also hinders the detection task. Figure 1.1 shows sample microscopy images displaying fluorescently labeled HIV-1 particles, where one can observe instances of the challenges involved in detecting fluorescent particles. The particle detection step may yield a set of detections that fails to include image regions corresponding to particles (missed detections). Similarly, the set of detections may include image regions corresponding to the cellular background or other non-particle regions (false detections). Using the set of detections, the goal of localization is to derive a set of *measurements* that explictly describe the image position of each particle. To track the fluorescent particles, spatial-temporal filtering is typically used. The aim of *spatial-temporal filtering* is to estimate the position of an object given a set of (position) measurements. This task is usually formulated within a Bayesian framework where the goal is to compute a posterior distribution on the position of the object conditioned on a set of position measurements. Challenges for this task entailed by fluorescence microscopy include the complex motion behavior of the particles, which includes abrupt changes in speed and direction, as well as the degree of uncertainty involved in the measurement process, which is typically related to the image noise. To obtain the corresponding (position) measurements of each object, associations between objects and position measurements are to be established. Finding these associations is the goal of *motion correspondence*. The task is simplified in cases where the objects and

the (position) measurements are sparsely located. Complications arise if the objects and the (positions) measurements are densely located, since the associations become ambiguous. Missed as well as false (position) measurements exacerbate the complexity of the correspondence task. In microscopy image sequences displaying fluorescent particles, typically images display a high object density, so the task of motion correspondence is not trivial. Assuming that the behavior of an individual particle can be decomposed into certain behavior labels, the aim of *behavior identification* is the derivation of these labels. Since the position of each particle is available, statistics (e.g., displacement, intensity) *representing* the behavior of an individual particle over time may be computed. The behavior identification task may be thus reformulated as estimating the behavior labels that give rise to the observed temporal representations. One issue is defining a set of labels that provides a suitable description of the considered behavior of a particle. Another issue is dealing with errors within the computed temporal representations. These errors may be caused by the nature of the microscopy images (e.g., the intensity over time of an individual particle is altered by photobleaching as well as by the image noise). Errors in the tasks of object localization and position estimation generally lead to errors of the temporal representations. In conclusion, the tasks involved in estimating dynamical behaviors of subcellular structures based on fluorescence microscopy image sequences are challenging.

1.3 Approach and Contributions

This dissertation is concerned with the estimation of dynamical behaviors of subcellular structures based on time-lapse fluorescence microscopy images. We deal with the three tasks entailed by this problem, namely object localization, tracking, as well as behavior identification. We focus on tracking, in particular, on spatial-temporal filtering, as well as on behavior identification, more specifically:

- To track individual fluorescent particles, we introduce an efficient and robust approach based on *probabilistic data association* (PDA). We propose a localization scheme that combines bottom-up and top-down strategies. The localization scheme is underpinned by an efficient *elliptical sampling scheme* based on the Gaussian densities underlying the spatial and temporal uncertainties involved in the tasks of particle localization and position estimation over time. The scheme explores systematically the solution space by generating *multiple samples* that are integrated via the principle of a *combined innovation* as outlined by the PDA algorithm. Since we use an elliptical sampling scheme we refer to our approach as PDAE. The approach includes a *recognition-by-synthesis* strategy that evaluates

an image likelihood which is constructed by directly querying the image data. Synthesis is performed via *parametric appearance models*, and we present models for round particles (e.g., HIV-1 particles) as well as for elongated particles (e.g., microtubule tips). *Multiple motion models* are also integrated in our tracking approach.

- To effectively exploit the image data during tracking and to cope with objects in close proximity, we construct an image likelihood. For efficiency, this likelihood is built independently for each tracked particle. In cases where multiple particles lie in close proximity, particles with high likelihoods may influence the position estimates of neighboring particles. To prevent this influence, we propose an iterative solution that computes the *image support* provided by each position to each tracked particle *relative* to the support given to neighboring particles. The relative image support takes into account the image likelihood of the particles as well as the spatial dependency between two image positions in general.

- To identify behaviors of individual fluorescent particles, we have developed an approach grounded within the theory of Bayesian estimation for identifying fusion of double labeled enveloped virus particles with the cell membrane in multi-channel fluorescence microscopy images. Our approach adopts a *layered* architecture that decomposes the actions of a single virus particle into three abstractions (viz., the intensity, the underlying temporal intensity model, and the behavior). The three abstractions are represented by the different layers, which are in turn described using a *stochastic hybrid system* and a *hidden Markov model* (HMM). The two representations are combined via a *maxbelief* strategy. The combination of stochastic hybrid systems and HMMs offer an improved modeling capability. To estimate the variables of the layers we use a *hybrid particle filter* as well as the *Viterbi algorithm*. We introduce non-linear autoregressive models to intuitively describe the behavior of the particles. The modularity of our layered approach endows it with efficiency and flexibility, and so the approach can also be straightforwardly adapted for identifying other behaviors.

- We have performed an extensive quantitative evaluation of deterministic and probabilistic tracking approaches based on real microscopy image sequences displaying HIV-1 particles. Such an evaluation is essential to predict the performance of the approaches under realistic conditions. The deterministic approaches are based on bottom-up localization schemes and a global nearest neighbor approach for motion correspondence. The probabilistic approaches are based on *Kalman filters*, a *mixture of particle filters* (MPF), our previous approach based on *independent*

particle filters (IPF) (e.g., [71], [70]), and our new approach based on probabilistic data association in conjunction with the elliptical sampling scheme (PDAE). The approaches have been applied to nine real microscopy image sequences (each comprising between 150 and 400 frames) displaying HIV-1 particles. We have compared the tracking results with ground truth obtained from manual tracking and quantified the performance.

- We have quantitatively studied the performance of our approaches for tracking as well as for behavior identification using synthetic 2D and 3D image sequences. We investigated the performance relative to the level of image noise which showed the robustness of our approaches to noisy conditions. We also analyzed the performance as a function of the object density. This also showed the scalability of our approach to a large number of objects. For behavior identification, we also studied the performance of our approach in image sequences where particles exhibit heterogeneous behaviors.

- We have applied our approaches to real 2D and 3D multi-channel fluorescence microscopy image sequences. The images display HIV-1 particles, herpes simplex virus (HSV) replication compartments or microtubule tips within neurons. This demonstrates the applicability of our approaches to a *broad range* of fluorescence microscopy image data.

- We successfully applied the PDAE approach to the 2D and 3D image data of the recent Particle Tracking Challenge at the International Symposium on Biomedical Imaging (ISBI) 2012, where 14 groups world-wide participated. The challenge comprised four different application scenarios (vesicles, virus particles, receptors, and microtubule tips) and different performance measures were computed. It turned out that our approach yielded the best overall performance (result announced at ISBI'2012).

- Our approaches for tracking and behavior identification have been used in biological applications in cooperation with our partners and have enabled novel biological findings. A main advantage in these applications is that our approaches yield scientific benefits in the form of objectivity, reproducibility, and statistical accuracy.

1.4 Overview of the Thesis

The thesis is organized as follows. Chapter 2 reviews prior work within the fields of computer vision as well as biological image analysis for the tasks of tracking as well as

behavior identification. In Chapter 3 we address the problem of tracking. We discuss probabilistic approaches and introduce our approach based on probabilistic data association. We also present models for tracking round objects as well as elongated objects. In Chapter 4 we tackle the problem of behavior identification using a layered probabilistic approach based on hybrid stochastic systems and hidden Markov models (HMMs). Models for the intensity over time as well as for the behavior of virus particles are introduced. In Chapter 5 we present experimental results for synthetic image sequences as well as for real microscopy image sequences displaying HIV-1 particles. We quantify the performance of the developed approaches and carry out an experimental comparison with previous approaches. We also present the results obtained for image sequences displaying HSV compartments as well as for image sequences displaying microtubule tips within a neuron's growth cone. Chapter 6 summarizes the results, presents our conclusions, and provides suggestions for future work.

Chapter 2

Overview of Previous Work

In this chapter, we review developments in the field of computer vision for tracking as well as for behavior identification. We focus on tracking approaches dealing with multiple objects. In particular, we discuss approaches for the task of tracking multiple fluorescent particles. Approaches for the task of behavior identification of fluorescent particles are also discussed.

2.1 Tracking Approaches in Computer Vision

Tracking multiple objects entails determining the position over time of each object in successive images of an image sequence. The task of tracking multiple objects may be solved in a *deterministic* way or in a *probabilistic* manner. We discuss both paradigms. We briefly elaborate also on approaches that post-process an initial set of trajectories (so-called *tracklets*) to obtain a refined set of trajectories. Since the latter type of approaches relies on an initial set of trajectories, the focus is on building such an initial set of trajectories given the image data only.

2.1.1 Deterministic Approaches

Deterministic approaches for tracking multiple objects typically follow a two-step paradigm encompassing the steps of *object localization* and *motion correspondence*. We briefly discuss object localization and focus on motion correspondence.

2.1.1.1 Object Localization

Determining the location of an object within an image entails two tasks: detection and localization. Depending on the type of object to be detected (e.g., pedestrians [157], [45], [63], [59], [66], faces [204], [216], eyes [77], cars [225], birds [49]), several different

approaches for object localization exist. Typically detection approaches combine the steps of image pre-processing, feature computation, and a classification step [183], [59]. Pre-processing may include a step for the normalization of the image intensities. For example, to detect faces [204], the intensities of an image are normalized using the image's intensity variance. If working with color images, pre-processing may also include a step for choosing an appropriate color space. For example, to detect birds, the RGB input images are mapped to a certain color space. The pre-processing stage may include the removal of the background (e.g., [49]). Once the images have been pre-processed, features (e.g., gradients) are calculated at each pixel. The features of several pixels may be aggregated over regions via histograms (e.g., histogram of the orientation of the gradients over a certain region [45]). The features or descriptors may be used as input for classification schemes (e.g., Adaboost-based classifiers [204]) that determine whether a candidate pixel or region corresponds to a true object. Once the regions that correspond to objects have been established, a *measurement* $\mathbf{y}$ may be computed for each region. Each measurement describes each object and may include information on its position, shape, and appearance. Thus, for each image, an object localization scheme yields a set $Y_t = \{\mathbf{y}_t^1, \mathbf{y}_t^2, \ldots, \mathbf{y}_t^{N_t}\}$ of N_t measurements representing the objects detected within the image at time step t. Since errors may arise in the steps involved in the task of object localization, the set Y_t may miss some true objects as well as include false objects. Also the computed position of each object may include a localization error.

2.1.1.2 Motion Correspondence

The task of *motion correspondence* entails determining whether two measurements $\mathbf{y}_t$ and $\mathbf{y}_{t'}$ obtained at different time steps t and t' originate from the same object [42]. In the case of deterministic approaches, this task typically amounts to determining the associations between sets of measurements Y_t and $Y_{t'}$ obtained at different time steps.

2.1.1.2.1 Correspondence over Two Time Steps To solve the problem of motion correspondence, one-to-one associations between measurements obtained at consecutive time steps t and $(t+1)$ may be established. Typically a heuristic (e.g., smoothness of motion) is translated into a distance function $d(\cdot, \cdot)$ that is used as a measure of the degree of correspondence between two measurements. A common strategy to determine correspondences is the *nearest neighbor* (NN) strategy, where a correspondence is established between a measurement $\mathbf{y}_t^i$ and the measurement $\mathbf{y}_{t+1}^j$ by:

$$\min_{\mathbf{y}_{t+1}^j \in Y_{t+1}} d(\mathbf{y}_t^i, \mathbf{y}_{t+1}^j). \tag{2.1}$$

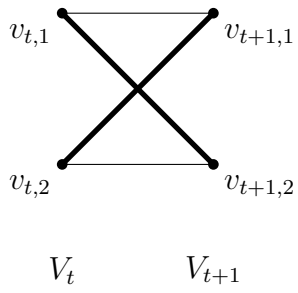


Figure 2.1: Sample bi-partite graph Γ with partitions V_t and V_{t+1}. Edges included in a possible matching M are drawn with a thick line.

To consider only the most likely associations, only associations entailing a distance $d(\cdot,\cdot)$ less than or equal to a certain threshold $d_{\max}$ are taken as valid. This amounts to searching for a nearest neighbor in a vicinity $V_{\mathbf{y}_t^i}(d_{\max})$ defined by $d_{\max}$ around $\mathbf{y}_t^i$:

$$V_{\mathbf{y}_t^i}(d_{\max}) \equiv \{\mathbf{y}_{t+1}^k \in Y_{t+1} \,|\, d(\mathbf{y}_t^i, \mathbf{y}_{t+1}^k) \leq d_{\max}\}. \tag{2.2}$$

The requirement presented in (2.1) can be reformulated as:

$$\min_{\mathbf{y}_{t+1}^k \in V_{\mathbf{y}_t^i}(d_{\max})} d(\mathbf{y}_t^i, \mathbf{y}_{t+1}^k). \tag{2.3}$$

For two different measurements $\mathbf{y}_t^i$ and $\mathbf{y}_t^k$, $i \neq k$ that are closely positioned and whose corresponding neighborhoods $V_{\mathbf{y}_t^i}(d_{\max})$ as well as $V_{\mathbf{y}_t^k}(d_{\max})$ overlap, the requirement specified in (2.3) may be fulfilled by the same measurement $\mathbf{y}_{t+1}^j$. In this case a *conflicting correspondence* arises. Since the NN strategy considers the correspondence of one measurement at time step t without taking into consideration the correspondences of the other measurements within the same time step, conflicting correspondences are not handled properly and this may lead to correspondence errors.

To take into consideration the correspondences of all measurements, the correspondence problem may be modeled using a weighted bi-partite graph $\Gamma = (V_t, V_{t+1}, E)$ where the vertices $V = \{V_t, V_{t+1}\}$ of the graph are given by the measurements: $v_{t,i} \equiv \mathbf{y}_t^i$. The two partitions V_t and V_{t+1} are induced by the two sets of measurements Y_t and Y_{t+1}. An edge $e_{ij} = \{v_{t,i}, v_{t+1,j}\} \in E$ between two vertices $v_{t,i}$ and $v_{t+1,j}$ stands for a possible correspondence between the measurements $\mathbf{y}_t^i$ and $\mathbf{y}_{t+1}^j$. A weight w_{ij} is assigned to each edge e_{ij} and is defined by the distance function $w_{ij} \equiv d(\mathbf{y}_t^i, \mathbf{y}_{t+1}^j)$. Assuming that only 1-to-1 correspondences are allowed, a set of valid correspondences for all measurements is represented by the set of edges with no shared vertices, i.e., a *matching* M. To denote

whether an edge e_{ij} is included in a matching M, we introduce an indicator variable a_{ij}:

$$a_{ij} = \begin{cases} 1 & e_{ij} \text{ is included in the matching } M \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

Since a correspondence must be established for each measurement, every vertex in the graph must be incident with an edge within the matching M. This introduces the following constraints:

$$\sum_{i=1}^{N_t} a_{ij} = 1, \qquad 1 \leq j \leq N_{t+1} \tag{2.5}$$

$$\sum_{j=1}^{N_{t+1}} a_{ij} = 1, \qquad 1 \leq i \leq N_t \tag{2.6}$$

Here we assume that the size of both partitions is the same, i.e., $N_t = N_{t+1}$. While this often does not hold in real applications, the situation may be remedied by adding 'dummy' measurements so that the equality is preserved (e.g., [199]). A sample bipartite graph and a possible matching is shown in Figure 2.1. Within this graph-theoretical setting, the task of finding correspondences between two sets of measurements Y_t and Y_{t+1} amounts to finding a complete matching M that minimizes the following *cost function* C_2 over two time steps:

$$C_2 = \sum_{i=0}^{N_t} \sum_{j=0}^{N_{t+1}} a_{ij}\, w_{ij}. \tag{2.7}$$

Minimizing this cost function C_2 can be achieved via optimization algorithms for the assignment problem, such as the Hungarian algorithm (e.g., [141], [199]) or via algorithms for the transportation problem (e.g., [47]). Because this approach takes into consideration the correspondences of all measurements, it is called the *global nearest neighbor* (GNN) approach (e.g., [19]). While the GNN strategy might deliver optimal results in terms of the cost function C_2, the correspondences may not be the 'best' in terms of correct trajectories. One reason for this is that the set of measurements at each time step may be missing some measurements (because of missed detections) or may include measurements that do not correspond to true objects (because of false detections). To resolve this, a larger number of time steps (images) may be used to establish the correspondences (e.g., [173], [94], [223], [14]).

2.1.1.3 Summary

In summary, deterministic approaches establish correspondences between the sets of position measurements generated by an independent localization scheme. Because the approaches rely on the detections provided by the localization scheme, the approaches are also known in the literature as *tracking-by-detection* approaches. Most deterministic approaches are thus restricted by the position measurements generated by an independent localization scheme. Also the localization error is not taken into consideration. To deal with correspondence conflicts, recent approaches use the measurements obtained over several time steps. The correspondence task is often formulated within the framework of graph theory and combinatorial optimization. Linear programming algorithms (e.g., Hungarian algorithm [141], simplex algorithm, min-cost flow algorithms [143], l-shortest paths) are often used to solve the optimization problems entailed by the graphs (e.g., bi-partite graphs, flow networks). In most approaches the distance function that measures the correspondence affinity is typically independent of previous time steps. Deterministic approaches also use relatively simple models for the motion of the objects. Finally, the localization step and the correspondence step are uncoupled. This entails that the localization scheme does not take into account the temporal information encoded in the image sequence, and similarly the correspondence step does not exploit all available image data. This diminishes the effectiveness of such approaches.

2.1.2 Probabilistic Approaches

Probabilistic approaches are characterized by defining a probability distribution on the variables describing the motion of the objects. These variables may include the position and velocity components of the object. In these cases, the probability distributions are typically estimated sequentially based on the image data using a spatial-temporal filter. The variables may also describe possible correspondences between the tracked objects and the detected position measurements. In these cases, the probability distributions may be computed using a probabilistic scheme for motion correspondence. Below, we review recent developments for spatial temporal filtering as well as probabilistic approaches for motion correspondence.

2.1.2.1 Spatial-Temporal Filtering

Within a Bayesian framework, the theory of sequential state estimation (e.g., [5]) assumes that an object is represented by a (hidden) state vector $\mathbf{x}_t$ and that a noisy measurement $\mathbf{y}_t$ reflects the true state $\mathbf{x}_t$. At time step t, the aim is to estimate the state $\mathbf{x}_t$ given a sequence of measurements $\mathbf{y}_{1:t}$. A Bayesian approach to the problem is

to estimate the posterior probability density function (pdf) $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, which represents a degree of "belief" on the state $\mathbf{x}_t$ given a series of measurements $\mathbf{y}_{1:t}$. To compute this posterior pdf, the recursive Bayesian filter ([55]) requires three elements: a dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, which represents the evolution of the state $\mathbf{x}_t$ over time, a measurement model $p(\mathbf{y}_t|\mathbf{x}_t)$, which establishes the relation between the state $\mathbf{x}_t$ and the measurement $\mathbf{y}_t$, and an initial prior $p(\mathbf{x}_0)$, which initializes the recursive computation. At each time step, the Bayesian filter computes recursively the posterior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ by first generating the prior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ using stochastic propagation (2.8) and then updating the predicted prior using Bayes' theorem (2.9):

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, d\mathbf{x}_{t-1} \tag{2.8}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{y}_{1:t-1}). \tag{2.9}$$

The recursive relation is often intractable since complex high-dimensional integrals must be solved. Notwithstanding, the Kalman filter provides an analytical solution if the dynamical and measurement models are assumed to be linear and Gaussian. If these assumptions do not hold, an Extended or Unscented Kalman filter may be used, yet the resulting posterior pdf remains Gaussian. If the posterior pdf is multimodal, parametric representations may be calculated via a Gaussian sum filter, which represents the posterior pdf with a mixture of Gaussians. A more general non-parametric approximation of the posterior pdf can be achieved via the particle filter. Below we review some of these spatial-temporal filter techniques.

2.1.2.1.1 Kalman Filter The Kalman filter (e.g., [5], [207]) represents the posterior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ via a Gaussian probability distribution. The Kalman filter is discussed in more detail in Chapter 3. Kalman filters have been used for tracking soccer players (e.g., [91], [211]), cars in aerial image sequences (e.g., [142]), faces, (e.g., [174]), pedestrians (e.g., [157], [4]), as well as bats (e.g., [209]). In these cases, one Kalman filter is instantiated for each object. Measurements obtained using an independent localization scheme are assigned to each filter using a motion correspondence algorithm (see Section 2.1.1.2). The Kalman filter is computationally efficient, since only some matrices need to be calculated and multiplied at each time step. Thus it scales well when tracking multiple objects. The main limitations of the Kalman filter are that: 1) only linear functions for the dynamical and measurement models are supported, and 2) the posterior pdf is assumed to be Gaussian, i.e., unimodal, whereas in certain applications this distribution may be multimodal. To address the first limitation, the Extended Kalman filter (EKF, e.g., [9]) as well as the Unscented Kalman filter (UKF, e.g., [96]) have been introduced. The second limitation may be dealt by using Gaussian sum filters (e.g., [5]) whereby

the posterior distribution is approximated by a mixture of Gaussian components. The approximation holds only for small values of the covariance matrices parametrizing the Gaussian posterior densities of the individual components. If these covariance values are too large, more general methods, such as the particle filter, are needed.

2.1.2.1.2 Particle Filter The particle filter (e.g., [72], [88]) is a numerical approach that has been quite successful for non-linear and/or non-Gaussian models. The key idea behind this approach is that the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is approximated with a set of random samples (the 'particles') that are associated with importance weights. More details on the particle filter are provided in Chapter 3. Particle filters have been used for tracking multiple objects. Existing approaches for tracking multiple objects can be classified into particle filters defined on a *multiple-body state space*, where the entire configuration of all objects is *jointly* estimated, and particle filters defined on a *one-body state space*, where the configuration of each object is *independently* estimated. The former approach, known as *joint particle filters* (e.g., [120], [89], [196], [100], [109], [181], [224], [15]), is quite suitable for modeling interactions and occlusions. The joint strategy is only applicable to few objects due to the computational effort that is involved in estimating the posterior distribution in the high-dimensional space entailed by the joint configuration of all objects. The latter approach, denoted as *independent particle filters* (e.g., [220], [67], [146], [214], [30], [27], [159]), while inducing lower computational requirements, require additional mechanisms (e.g., using an inertia model that influences the likelihood of each sample [146]) to prevent the independent filters from latching to the object with the best likelihood ([100]).

2.1.2.1.3 Rao-Blackwellized Particle Filter When estimating a state vector $\mathbf{x}_t$ in a high dimensional state space $\mathcal{X}$, a large number of samples is required to thoroughly explore this space. Assuming that the state space $\mathcal{X}$ can be partitioned into two subspaces $\mathcal{R}$ and $\mathcal{B}$, the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = p(\mathbf{r}_t, \mathbf{b}_t|\mathbf{y}_{1:t}),\ \mathbf{r}_t \in \mathcal{R},\ \mathbf{b}_t \in \mathcal{B}$ conditioned on a sequence of measurements $\mathbf{y}_{1:t}$ can be written via Bayes' theorem and stochastic propagation as:

$$p(\mathbf{r}_t, \mathbf{b}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{r}_t, \mathbf{b}_t)p(\mathbf{r}_t, \mathbf{b}_t|\mathbf{y}_{1:t-1}). \tag{2.10}$$

Approximating this posterior with the standard particle filter would require sampling jointly for $\mathbf{r}_t$ and $\mathbf{b}_t$. Besides the computational burden that this entails, over time the distribution of the weights of the samples in this large state space tends to be very skewed (i.e., only few samples would have a large weight, while several samples would tend to have much lower weights). This also reduces the accuracy of the sample-based

approximation [56].

One strategy to cope with the large dimension of the state space is to marginalize out (i.e., integrate over) some variables from the posterior [129], [99]. Integrating over $\mathbf{b}_t$, the marginal $p(\mathbf{r}_t|\mathbf{y}_{1:t})$ can be obtained by:

$$p(\mathbf{r}_t|\mathbf{y}_{1:t}) \propto \int p(\mathbf{y}_t|\mathbf{r}_t, \mathbf{b}_t) p(\mathbf{r}_t, \mathbf{b}_t|\mathbf{y}_{1:t-1}) d\mathbf{b}_t. \tag{2.11}$$

Within this strategy, the assumption is that the posterior $p(\mathbf{r}_{t-1}, \mathbf{b}_{t-1}|\mathbf{y}_{1:t-1})$ can be factored via the chain rule of probability as:

$$p(\mathbf{r}_{t-1}, \mathbf{b}_{t-1}|\mathbf{y}_{1:t-1}) = p(\mathbf{b}_{t-1}|\mathbf{r}_{t-1}, \mathbf{y}_{1:t-1}) p(\mathbf{r}_{t-1}|\mathbf{y}_{1:t-1}), \tag{2.12}$$

Assuming that $\mathbf{b}_{t-1}$ is independent of $\mathbf{r}_t$ given $\mathbf{r}_{t-1}$ and also that $\mathbf{b}_t$ is independent of $\mathbf{y}_{1:t-1}$ given $\mathbf{b}_{t-1}$, and after some manipulations, the marginal $p(\mathbf{r}_t|\mathbf{y}_{1:t})$ (2.11) can be rewritten as:

$$p(\mathbf{r}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{r}_t|\mathbf{y}_{1:t-1}) \int p(\mathbf{y}_t|\mathbf{r}_t, \mathbf{b}_t) \left[\int p(\mathbf{b}_t|\mathbf{b}_{t-1}, \mathbf{r}_t) p(\mathbf{b}_{t-1}|\mathbf{r}_{t-1}, \mathbf{y}_{1:t-1}) d\mathbf{b}_{t-1} \right] d\mathbf{b}_t. \tag{2.13}$$

The terms inside the outer integral in (2.13) can be seen as proportional to the posterior density $p(\mathbf{b}_t|\mathbf{r}_t, \mathbf{y}_{1:t})$ conditioned on *both* the sequence of measurements $\mathbf{y}_{1:t}$ as well as on $\mathbf{r}_t$. If this density can be computed analytically, then only samples for $\mathbf{r}_t$ are needed. Since the dimension of $\mathcal{R}$ is smaller than that of the joint state space, fewer samples are required [129]. Within the context of Bayesian filtering, the marginalization strategy is known as 'Rao-Blackwellization'. A particle filter used for solving the recursion in (2.13) is thus known as the Rao-Blackwellized particle filter. The details of the Rao-Blackwellized particle filter are given in [129].

While a smaller state space is explored, the Rao-Blackwellized particle filter estimates a Gaussian distribution for *each* sample. This is computationally expensive because for each sample a parametric representation must be updated over time. Certainly, samples are proposed in a smaller space, so the accuracy of the sample-based representation is better. Nonetheless, the assumptions for the Rao-Blackwellized particle filter are relatively demanding: it assumes that a one-way dependency between the variables can be established, and also that the dynamical and measurement models of some of the variables are linear. Thus the Rao-Blackwellized particle filter may not be generally applicable. Nevertheless it has been used for tracking single objects (e.g., single bees [99] as well as single pedestrians [212]).

2.1.2.1.4 Summary We have reviewed recent approaches for estimating the state vector $\mathbf{x}_t$ given a sequence of measurements $\mathbf{y}_{1:t}$. By modeling the temporal evolution using a dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and incorporating measurements derived from the images via a measurement model $p(\mathbf{y}_t|\mathbf{x}_t)$, a spatial-temporal filter estimates the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ from which an estimate for $\mathbf{x}_t$ can be obtained. For linear and Gaussian models, the Kalman filter provides a fast and exact solution. For non-linear models, Extended Kalman filters as well as Unscented Kalman filters are applicable, yet the posterior pdf $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ remains Gaussian. Particle filters provide a non-parametric approach to deal with non-linear, non-Gaussian models by using a sample-based representation of the posterior pdf. If certain elements of the state vector $\mathbf{x}_t$ can be analytically estimated, the state space can be more efficiently explored using a Rao-Blackwellized particle filter that proposes samples in a reduced state space.

For tracking multiple objects, Kalman filters (and related approaches, e.g., the Extended Kalman filter) remain popular because of their low computational demands. Approaches based on the Kalman filter incorporate only the measurements computed by an independent localization algorithm; thus localization and position estimation are uncoupled steps. In this sense, approaches based on Kalman filters may be considered as tracking-by-detection approaches, too. In contrast, the particle filter queries directly multiple image positions (encoded by the position elements of the random samples) to determine the location of an object. The filter thus combines localization and position estimation and achieves more robust results. Because of the random nature of its *top-down* localization scheme, the filter uses a relatively large number of samples. This introduces a high computational cost because an image likelihood is evaluated for each sample. While one can improve the efficiency of the particle filter by, for example, marginalization (cf. Section 2.1.2.1.3), several hundred samples are still required to ensure a good approximation of the posterior pdf.

2.1.2.2 Motion Correspondence

Probabilistic approaches for motion correspondence evaluate the probability of possible correspondences between predicted measurements and the measurements obtained via an independent localization scheme. These approaches are generally characterized by using a spatial-temporal filter for building the trajectories. A review of probabilistic approaches is provided in [42]. Here we briefly discuss an approach determining correspondences over two time steps, namely the joint probabilistic data association (JPDA) algorithm. We also discuss correspondence approaches taking into account several time steps. We consider the multiple-hypothesis tracking (MHT) algorithm as well as approaches based on sampling schemes (e.g., Markov Chain Monte Carlo (MCMC)

methods).

2.1.2.2.1 Joint Probabilistic Data Association To determine the correspondences between a set $\hat{Y}_t$ of H predicted measurements as computed by a set of spatial-temporal filters tracking H objects, and a set Y_t consisting of N_t measurements obtained by an independent localization scheme, one could use a global nearest neighbor scheme (GNN; see Section 2.1.1.2.1) where the distance function $d(\cdot,\cdot)$ used to quantify the degree of correspondence is given by the *Mahalanobis distance*:

$$d_{\mathbf{S}^h}(\mathbf{y}^i, \hat{\mathbf{y}}^h) = (\mathbf{y}^i - \hat{\mathbf{y}}^h)^T \mathbf{S}^{h^{-1}} (\mathbf{y}^i - \hat{\mathbf{y}}^h) \tag{2.14}$$

entailed by the associated measurement $\mathbf{y}^i$ and the predicted measurement $\hat{\mathbf{y}}^h$ and predicted measurement covariance matrix $\mathbf{S}^h$ (see (3.12)) as calculated by each h-th spatial-temporal filter; for readability, we have omitted the time index t of the measurements. The GNN yields as a solution an assignment matrix $\mathbf{A}$ where each element a_{ih} indicates whether a measurement i is associated with a spatial-temporal filter h. Since the set Y_t may not include measurements for some of the H tracked objects and may also include a certain number N_{FD} of false detections, the assignment $\mathbf{A}$ determined by the GNN approach might lead to tracking errors. Instead of relying on a single (possibly incorrect) assignment $\mathbf{A}$, a more robust approach would take into consideration all possible assignments. When considering multiple assignments, each r-th possible assignment $\mathbf{A}^r$ establishes a correspondence between a spatial-temporal filter h and a single measurement. Multiple assignments thus imply that a single spatial-temporal filter h may be associated with multiple measurements. Since not all measurements may entail the same degree of correspondence relative to a given h spatial-temporal filter, a *probability of association* β_{ih} is computed for each i-th measurement. To update a single h spatial-temporal filter using multiple measurements, a combined innovation $\boldsymbol{\nu}^h$:

$$\boldsymbol{\nu}^h = \sum_{i=1}^{N_t} \beta_{ih} (\mathbf{y}^i - \hat{\mathbf{y}}^h) \tag{2.15}$$

is computed. The combined innovation $\boldsymbol{\nu}^h$ is the inherent idea of the *probabilistic data association* (PDA) algorithm. In the PDA algorithm, the association probability β_{ih} is defined via the posterior probability $P(a_{ih}|Y_t)$ of the association a_{ih} between the measurement $\mathbf{y}_i$ and the filter h. This posterior probability is only useful in cases where only single objects are tracked, since it does not take into account the probability of the associations a_{jl}, $j \neq i, l \neq h$ between other measurements and other filters. In comparison, the joint probabilistic data association (JPDA) algorithm takes into consideration all

possible assignments between filters and measurements. Instead of using only a single assignment $\mathbf{A}$, the JPDA algorithm enumerates all feasible assignments $\mathbf{A}^r$. In this case the association probability β_{ih} between a measurement i and the spatial-temporal filter h is relative to the posterior probability $P(\mathbf{A}^r|Y_{1:t})$ of each assignment $\mathbf{A}^r$ where the association a^r_{ih} has been established, i.e., $a^r_{ih} = 1$:

$$\beta_{ih} \equiv \sum_{\mathbf{A}^r} P(\mathbf{A}^r|Y_{1:t}) a^r_{ih}. \tag{2.16}$$

The posterior probability $P(\mathbf{A}^r|Y_{1:t})$ is given by Bayes' theorem [42] as:

$$P(\mathbf{A}^r|Y_{1:t}) \propto p(Y_t|\mathbf{A}^r, Y_{1:t-1}) P(\mathbf{A}^r) \tag{2.17}$$

where $\mathbf{A}^r$ is assumed to be independent of the previous sets of measurements $Y_{1:t-1}$. To compute the likelihood of the current measurements Y_t given the assignment $\mathbf{A}^r$, the approach assumes that $p(Y_t|\mathbf{A}^r, Y_{1:t-1})$ can be decomposed over all N_t individual measurements $\mathbf{y}^j$:

$$p(Y_t|\mathbf{A}^r, Y_{1:t-1}) = \prod_j^{N_t} p(\mathbf{y}^j|\mathbf{A}^r, Y_{1:t-1}) \tag{2.18}$$

If the assignment $\mathbf{A}^r$ establishes that the measurement $\mathbf{y}^j$ is associated with a certain spatial-temporal filter m, then the likelihood for the measurement $p(\mathbf{y}^j|\mathbf{A}^r, Y_{1:t-1})$ can be evaluated using the predicted measurement density $\mathcal{N}(\mathbf{y}^j; \hat{\mathbf{y}}^m, \mathbf{S}^m)$ (see Section 3.5) of the associated filter. Here $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. However, if the assignment $\mathbf{A}^r$ establishes that the measurement $\mathbf{y}^j$ is not associated with any filter, then the approach assumes that it is a false detection. The probability of false detection is assumed to follow a uniform distribution over the entire field of view V. Thus:

$$p(\mathbf{y}^j|\mathbf{A}^r, Y_{1:t-1}) = \begin{cases} \mathcal{N}(\mathbf{y}^j; \hat{\mathbf{y}}^m, \mathbf{S}^m) & a^r_{jm} = 1 \\ 1/V & \mathbf{y}^j \text{ is not associated with any object} \end{cases} \tag{2.19}$$

Note that some of the measurements that are not associated with any object (second case in (2.19)) may represent new objects in the field of view, but the standard JPDA algorithm does not distinguish between false detections and new objects.

To evaluate the prior probability $P(\mathbf{A}^r)$ of an assignment $\mathbf{A}^r$, the JPDA algorithm takes into account the number of assigned measurements N_{assigned} and the number of false detections $N_{\text{FD}} = N_t - N_{\text{assigned}}$, i.e., $N_t = N_{\text{assigned}} + N_{\text{FD}}$. The probability for the number of assigned measurements N_{assigned} is defined as the probability of detecting N_{assigned} true measurements out of all H currently tracked objects. This probability

follows the binomial distribution:

$$P(N_{\text{assigned}}|H) = \binom{H}{N_{\text{assigned}}} P_D^{N_{\text{assigned}}} (1 - P_D)^{H - N_{\text{assigned}}}, \tag{2.20}$$

where P_D is the probability of detecting a single object and $\binom{H}{N_{\text{assigned}}} = \frac{H!}{N_{\text{assigned}}!(H-N_{\text{assigned}})!}$ is the binomial coefficient that indicates the number of ways of choosing N_{assigned} measurements out of H potential measurements for all H currently tracked objects. The probability $P(N_{\text{FD}})$ for the number false detections N_{FD} follows either a Poisson distribution or a uniform distribution. Choosing a Poisson distribution leads to the parametric version of the JPDA algorithm while selecting a uniform distribution leads to the non-parametric version of the JPDA algorithm [42]. So far only the probabilities of the numbers $(N_{\text{assigned}}, N_{\text{FD}})$ alone have been considered. The assignment $\mathbf{A}^r$ also entails a specific way (i.e., arrangement) of choosing N_{assigned} as well as N_{FD} measurements out of all N_t measurements. The number of ways of choosing specifically N_{assigned} measurements as well as N_{FD} measurements out of N_t measurements is given by:

$$\binom{N_t}{N_{\text{assigned}}} \binom{N_t - N_{\text{assigned}}}{N_{\text{FD}}}, \tag{2.21}$$

and so assuming that each combination is equally likely, the probability of a specific combination is [152]:

$$P(\text{combination}|N_{\text{assigned}}, N_{\text{FD}}) = \frac{1}{\binom{N_t}{N_{\text{assigned}}}}. \tag{2.22}$$

Finally, $\mathbf{A}^r$ also specifies exactly a way of assigning N_{assigned} measurement to H objects. The number of ways in which N_{assigned} may be assigned to H objects is given by the following permutation:

$$\frac{H!}{(H - N_{\text{assigned}})!} \tag{2.23}$$

so the probability of each assignment given a certain combination is:

$$P(\text{assignment}|\text{combination}) = \frac{(H - N_{\text{assigned}})!}{H!}. \tag{2.24}$$

Putting together all these probabilities and after some simplifications one obtains the prior probability $P(\mathbf{A}^r)$:

$$P(\mathbf{A}^r) \propto \frac{N_{\text{FD}}!}{N_t!} P(N_{\text{FD}}) P_D^{N_{\text{assigned}}} (1 - P_D)^{H - N_{\text{assigned}}}. \tag{2.25}$$

In comparison to deterministic approaches for finding correspondences over two time

steps (see Section 2.1.1.2.1), the JPDA algorithm considers all possible assignments between the measurements and the objects. This amounts to considering associations between all measurements and all objects. By evaluating the likelihood of each feasible (global) assignment $\mathbf{A}^r$, the JPDA algorithm weights the reliability of a particular association between a measurement and an object. Since the reliability of each assignment is taken into consideration, the JPDA algorithm leads to better results in comparison to deterministic approaches, which essentially assign a reliability of one to the selected solution and zero to all other possible solutions. The robustness of the JPDA approach, however, incurs a strong degradation in the computational performance because all possible assignments must be enumerated and evaluated. The JPDA also assumes that the number of objects is constant over time, although this has been addressed by, for instance, assuming that the number of objects entering the field of view as well as the number of objects leaving the field of view follow a Poisson distribution [167], [50]. The JPDA algorithm has been used for tracking a small number of objects in computer vision. For example, the approach in [149] approximates the JPDAF posterior $P(\mathbf{A}^r|Y_{1:t})$ (see 2.17) with the likelihood $p(\mathbf{z}_t|\mathbf{X}_t)$ of an image $\mathbf{z}_t$ given a particular joint configuration $\mathbf{X}_t$ of all objects, which can be construed as an assignment between the predicted positions of the objects and the corresponding image positions. The approach is similar to approaches using joint particles filters (e.g., [120]) that also evaluate an image likelihood given a certain joint configuration ('sample'). The approach in [149] exploits effectively the image data. This also increases the computational demands of the approach since the joint image likelihood $p(\mathbf{z}|\mathbf{X}_t)$ has to be evaluated for each possible joint configuration $\mathbf{X}_t$. This is computational expensive because, to cope with occlusions, the approach lists all permutations of the *depth order* of all objects, which leads to a rather large number of possible configurations $\mathbf{X}_t$, even for a relatively small number of objects. In the standard JPDAF as well as in [149], Kalman filters are used to estimate the position of the individual objects. Recently, particle filters defined on a one-body state space have been used instead (e.g., [167]). In this case the likelihood $p(\mathbf{y}^j|\mathbf{A}^r, Y_{1:t-1})$ for a measurement $\mathbf{y}^j$ assigned with a spatial-temporal filter m (first case in (2.19)) is given by the mean likelihood obtained over all the samples of the $m-th$ particle filter. Since the image likelihood $p(\mathbf{y}_t|\mathbf{x}_t)$ of each sample needs to be evaluated, the approach also entails a large computational burden. For this reason the approach is only applicable to a small number of objects (e.g., up to six objects [29]).

2.1.2.2.2 Multiple-Hypothesis Tracking Analogous to the deterministic case, probabilistic approaches for motion correspondence benefit from taking into consideration measurements obtained over several time steps. One assumption used by deterministic approaches finding correspondences over several time steps is that the distance

function $d(\cdot,\cdot)$ that quantifies the degree of correspondence between two measurements is independent of previous time steps. In cases where a spatial-temporal filter is used to estimate the position, the trajectories are built recursively so correspondences established in the previous $t-1$ time steps influence the possible correspondences in the current time step t. In other words, previous and current correspondences are not independent, and so assuming that the distance function $d(\cdot,\cdot)$ is independent of previous correspondences is not very suitable.

One way to take into account previous correspondences is to evaluate all possible assignments between time steps $t-1$ and t for a set of disjoint trajectories $\mathcal{A}_{t-1}$. For each $\mathcal{A}_{t-1}^i$, each possible assignment $\mathbf{A}_t^j$ between the trajectories $\mathcal{A}_{t-1}^i$ and the set of measurements Y_t leads to a set of trajectories $\mathcal{A}_t^j$. In other words, each $\mathcal{A}_{t-1}^i$ branches into multiple $\mathcal{A}_t^j$. Over time, this leads to a tree of possible disjoint trajectories, where each leaf node of the tree essentially represents a *hypothesis* $\mathcal{A}_t^i$ about a possible set of trajectories. Since maintaining the full hypotheses tree leads to an exponential increase in computational effort over time, the posterior probability $P(\mathcal{A}_t^j|Y_{1:t})$ of each set of trajectories $\mathcal{A}_t^j$ given a sequence of sets of measurements $Y_{1:t}$ is evaluated. Hypotheses with a low probability are discarded thereby pruning the hypotheses tree. Since several hypotheses are considered, this approach is called the Multiple-Hypothesis Tracking (MHT) algorithm [152], [19], [42]. Since each hypothesis $\mathcal{A}_t^j$ is defined via its parent hypothesis $\mathcal{A}_{t-1}^i$ together with the assignment $\mathbf{A}_t^j$ leading to $\mathcal{A}_t^j$, the probability $P(\mathcal{A}_t^j|Y_{1:t})$ can be formulated via Bayes' theorem as:

$$\begin{aligned} P(\mathcal{A}_t^j|Y_{1:t}) &= P(\mathbf{A}_t^j, \mathcal{A}_{t-1}^i|Y_t, Y_{1:t-1}) && (2.26)\\ &\propto p(Y_t|\mathbf{A}_t^j, \mathcal{A}_{t-1}^i, Y_{1:t-1})P(\mathbf{A}_t^j|\mathcal{A}_{t-1}^i, Y_{1:t-1})P(\mathcal{A}_{t-1}^i|Y_{1:t-1}). && (2.27)\end{aligned}$$

This implicates evaluating the conditional likelihood of the current set of measurements $p(Y_t|\mathbf{A}_t^j, \mathcal{A}_{t-1}^i, Y_{1:t-1})$ as well as the prior probability on the current assignment $P(\mathbf{A}_t^j|\mathcal{A}_{t-1}^i, Y_{1:t-1})$. The exact formulas for these terms are given in [152].

In comparison to deterministic approaches finding correspondences over several time steps (e.g., [173], [94], [223], [14]), the distance function used by the MHT depends on previous correspondences. This advantage, however, comes at the expense of having to maintain a large number of hypotheses, which entails updating a large number of trajectories. Also, the MHT relies explicitly on spatial-temporal filters to estimate the positions, whereas deterministic schemes use simplified motion and measurement models that do not encode the uncertainty of the motion and measurement, respectively. However, using spatial temporal filters also introduces additional computational demands. For these reasons the MHT is usually considered to be quite computationally demanding. Another drawback of the standard MHT is that it does not take into con-

sideration directly the image data. In other words, the image information is not directly used to evaluate the involved probabilities. Nonetheless, in computer vision, the MHT algorithm has been used to track corners (e.g., [43]), vehicles (e.g., [185]) as well as the components of articulated objects (e.g., [114], [34]).

2.1.2.2.3 Approaches Based on Sampling Schemes The underlying idea of the MHT algorithm is to explore the space of possible disjoint trajectories over a certain time period. This is done by deterministically enumerating all possible sets of disjoint trajectories and evaluating their probability via the corresponding posterior (see (2.27)). Another way of exploring the space of possible disjoint trajectories is by drawing random samples from a distribution defined on this space. This is the main idea of approaches for motion correspondence based on sampling schemes (e.g., Markov Chain Monte Carlo methods, MCMC [73]). Here, one can distinguish between approaches drawing samples in an iterative manner (e.g., via the Metropolis-Hastings algorithm [79]) and approaches drawing samples in a sequential manner (e.g., via a Rao-Blackwell particle filter [168], [163]).

Approaches based on Markov Chain Monte Carlo methods (MCMC, e.g., [135], [219], [25]) draw samples from the following posterior distribution:

$$P(\mathcal{A}_T|Y_{1:T}) \propto P(Y_{1:T}|\mathcal{A}_T)P(\mathcal{A}_T). \tag{2.28}$$

As opposed to the MHT algorithm, the set of disjoint trajectories $\mathcal{A}_T$ takes into consideration all time steps T comprising the evaluated image sequence. In this case, the Metropolis-Hastings (MH) algorithm is used to obtain samples from (2.28). The idea of the MH algorithm is to obtain a set of samples by iterating along a Markov chain such that its *stationary distribution* is set to $P(\mathcal{A}_T|Y_{1:T})$, i.e., such that the samples generated by the Markov chain are distributed according to $P(\mathcal{A}_T|Y_{1:T})$. The iteration is driven by a proposal distribution $q(\mathcal{A}_T, \mathcal{A}'_T)$ that proposes a sample $\mathcal{A}'_T$ taking into account the current sample $\mathcal{A}_T$. The proposed sample $\mathcal{A}'_T$ is *accepted* as the next sample with acceptance probability $\alpha(\mathcal{A}_T, \mathcal{A}'_T)$:

$$\alpha(\mathcal{A}_T, \mathcal{A}'_T) = \min\left(1, \frac{P(\mathcal{A}'_T|Y_{1:T})q(\mathcal{A}'_T, \mathcal{A}_T)}{P(\mathcal{A}_T|Y_{1:T})q(\mathcal{A}_T, \mathcal{A}'_T)}\right). \tag{2.29}$$

If the sample is not accepted, then the sampling scheme stays at $\mathcal{A}_T$. To generate a sample $\mathcal{A}'_T$, the proposal $q(\mathcal{A}_T, \mathcal{A}'_T)$ may choose a certain *move* (i.e., an update of $\mathcal{A}_T$ that leads to a particular $\mathcal{A}'_T$). In [135], the proposal consists of the following moves (1) initialization and termination of trajectories, (2) splitting and merging of trajectories, (3) extension and reduction of trajectories, (4) update of trajectories, and

(5) trajectory switching. Using these moves to expand a set of empty trajectories would be very time consuming. Instead, the MH sampler is initialized with a set of disjoint trajectories obtained with a sequential approach based on Kalman filters and a nearest neighbor scheme for motion correspondence. In [135], the set of position measurements is given by an independent detection scheme. In [219], moves that generate new position measurements by analyzing the image data are also proposed and so the approach exploits more effectively the image data. Typically, trajectories are modeled via an auto-regressive process embedded into a Kalman filter. In [25], trajectories are instead modeled using Gaussian Processes (GPs) (e.g., [150]). Using GPs for generating the trajectories leads to smoother trajectories. One disadvantage of approaches based on MCMC methods for motion correspondence is that they do not support a sequential analysis of the image data. In other words, all time steps must be available to apply the approach.

Approaches based on Rao-Blackwellized particle filters (e.g., [168]) instead analyze sequentially the image sequence to build the trajectories. Here the approach estimates both the joint configuration $\mathbf{X}_t$ of all objects as well as the set of disjoint trajectories $\mathcal{A}_{1:t}$, i.e., the concatenation of assignments $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_t$. By conditioning the joint configuration $\mathbf{X}_t$ on $\mathcal{A}_{1:t}$, the posterior distribution can be factored in the following way:

$$p(\mathbf{X}_t, \mathcal{A}_{1:t}|Y_{1:t}) = p(\mathbf{X}_t|\mathcal{A}_{1:t}, Y_{1:t})P(\mathcal{A}_{1:t}|Y_{1:t}). \tag{2.30}$$

This posterior is amenable to the Rao-Blackwellized particle filter (see Section 2.1.2.1.3), where the idea is to estimate $P(\mathcal{A}_{1:t}|Y_{1:t})$ by sampling, and to determine analytically $p(\mathbf{X}_t|\mathcal{A}_{1:t}, Y_{1:t})$ based on the samples obtained for $P(\mathcal{A}_{1:t}|Y_{1:t})$. To obtain samples from $P(\mathcal{A}_{1:t}|Y_{1:t}) \propto P(\mathbf{A}_t|\mathcal{A}_{1:t-1}, Y_{1:t})P(\mathcal{A}_{1:t-1}|Y_{1:t-1})$, the MH sampling scheme is used to generate samples from $P(\mathbf{A}_t|\mathcal{A}_{1:t-1}, Y_{1:t})$. Note that here the Markov chain explores only the space of assignments between two consecutive time steps and so in a sense the approach is similar to the MHT algorithm, where the assignments are instead enumerated in a deterministic way. Re-sampling the set of random samples to preserve the most likely samples is also analogous to pruning the hypotheses tree in the MHT algorithm.

While the MCMC approach in principle explores the entire space of sets of disjoint trajectories, the approach reduces the complexity entailed by the exploration of this space by obtaining an initial set of trajectories via an alternative tracking approach (e.g., based on Kalman filters and the GNN approach for motion correspondence). This allows application of the approach to a relatively large number of objects (e.g., up to 100 objects [135]). The approach may be also applicable in a sequential way by applying it only to a temporal window. The approach based on the Rao-Blackwellized particle filter builds the trajectories in a sequential manner that resembles the MHT algorithm.

As opposed to the latter, not all assignments (i.e., hypotheses) are listed, so it is more efficient. For both approaches, obtaining a single set of disjoint trajectories $\mathcal{A}_T$ as an output for the approach is not so straightforward because the computed distribution may be multimodal, i.e., multiple sets of of disjoint trajectories may provide a suitable solution for the correspondence problem.

2.1.2.2.4 Summary This section presented probabilistic approaches for finding correspondences. Approaches may take into consideration correspondences over two time steps (e.g., approaches based on the JPDA algorithm), over a larger number of time steps (e.g., approaches based on the MHT algorithm) as well as over all time steps (e.g., approaches based on MCMC methods). While the MHT algorithm in principle takes more time steps into account, it has been empirically shown [102] that the performance of the PDA algorithm together with the interacting multiple model (IMM) filter that accounts for multiple modes of motion is comparable to that of the MHT algorithm. This suggests that two images that are temporally distant may not contribute towards explaining the phenomena shown in either image. Approaches based on the Rao-Blackwellized particle filter are reported to perform comparatively well relative to the MHT algorithm [168]. The MCMC approach outperforms the MHT algorithm [135]. Certainly, the approach based MCMC requires the results of an additional tracking approach to reduce the solution space.

In comparison to deterministic approaches, probabilistic approaches for motion correspondence use spatial-temporal filters to build the trajectories. Since the spatial-temporal filters allow computing a predicted measurement density $\mathcal{N}(\cdot; \hat{\mathbf{y}}^h, \mathbf{S}^h)$, the Mahalanobis distance $d_{\mathbf{S}^h}(\mathbf{y}^i, \hat{\mathbf{y}}^h)$ may be used to evaluate the likelihood of association between a measurement $\mathbf{y}^i$ and a predicted measurement $\hat{\mathbf{y}}^h$. By virtue of using the predicted measurement covariance $\mathbf{S}^h$, the Mahalanobis distance takes into account the uncertainty that has accumulated over time by choosing particular assignments over time. In contrast to deterministic schemes for finding correspondences, the distance function used by probabilistic approaches depends on previous time steps. Also because spatial-temporal filters are used, trajectories are built in a more robust way. For example, in time steps where measurements may be missing, the predicted measurement may be used to guide the estimation process. Certainly, probabilistic approaches are more computationally demanding than deterministic since often the space of assignments over time needs to be explored and this may hamper their application in scenarios that involve a large number of objects.

2.1.3 Approaches Based on Tracklets

Given an initial set of trajectories (i.e., tracklets), the idea is to find a more accurate set of trajectories by modifying the initial set. Approaches based on tracklets (e.g., [184], [86]) often consider the task of finding tracklets that originated from the same object, i.e., finding correspondences among broken trajectories. To find correspondences among tracklets, the Hungarian algorithm is used. Approaches (e.g., [142]) also consider cases where individual tracklets may merge into a (joint) tracklet or where a (joint) tracklet may split into individual tracklets. In this case, correspondences between the joint and individual tracklets are also found via the Hungarian algorithm. In this case, finding correspondences between individual tracklets and a joint tracklet involves finding correspondences between an arbitrary number of individual tracklets and a single joint tracklet. However, the Hungarian algorithm supports only one-to-one correspondences. As such, the individual tracklets are represented via a single abstract tracklet that subsumes the individual tracklets. The problem of finding correspondences between tracklets has been also addressed by formulating the problem using a flow network (e.g., [209]). We refrain from delving further into these type of approaches since in any case these approaches rely on an initial set of trajectories. An accurate set of initial trajectories leads to better results for these approaches, so our focus is on obtaining reliable trajectories from the image data without any preliminary tracking step.

2.2 Tracking Approaches for Fluorescent Particles

In this section, we review approaches for tracking fluorescent biological particles in microscopy images. Analogous to the section on tracking approaches in computer vision, we discuss deterministic approaches as well as probabilistic approaches.

2.2.1 Deterministic Approaches

Deterministic approaches typically address only the tasks of particle localization and motion correspondence. We review schemes for particle localization as well as schemes for motion correspondence.

2.2.1.1 Particle Localization

2.2.1.1.1 Bottom-up Localization Schemes To address the problem of particle localization, either a bottom-up or a top-down strategy is generally adopted. Bottom-up approaches are driven by the image data, while top-down approaches are guided by prior knowledge on the expected shape and appearance of particles. Bottom-up localization

schemes for fluorescent particles typically comprise three consecutive steps: image pre-processing, particle detection, and particle localization. In the *pre-processing* step, the aim is threefold: first, to enhance the image intensities corresponding to particles; second, to suppress the image noise; and third, to attenuate the intensities of background structures (e.g., cellular background). One approach to pre-process the images is to use linear filters (e.g., [166]). One common filter is the Gaussian filter. Given a 2D image $g(x, y)$, the filtered image $g_{\text{Gauss}}(x, y)$ is described by the following convolution:

$$g_{\text{Gauss}}(x, y) = G^{\text{2D}}_{\sigma_{\text{F},xy}} * g(x, y) \tag{2.31}$$

$$= \sum_{i=-n}^{n} \sum_{j=-n}^{n} G^{\text{2D}}_{\sigma_{\text{F},xy}}(i, j) g(x - i, y - j) \tag{2.32}$$

where the width N_G of the Gaussian kernel $G^{\text{2D}}_{\sigma_{\text{F},xy}}$ is given by $N_G = 2n + 1$ and the kernel itself is defined by:

$$G^{\text{2D}}_{\sigma_{\text{F},xy}}(x, y) = \frac{1}{2\pi\sigma^2_{\text{F},xy}} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2_{\text{F},xy}}\right). \tag{2.33}$$

Doing a convolution with the kernel in (2.33) may be also seen as performing a correlation with such a kernel. In that sense, image regions that resemble the Gaussian kernel are enhanced. Since the appearance of sub-resolution fluorescent particles resembles a Gaussian function, image regions corresponding to fluorescent particles are enhanced. Another filter used for pre-processing the images is the Laplacian-of-Gaussian (LoG) filter. In this case, the kernel $\text{LoG}^{\text{2D}}_{\sigma_{\text{F},xy}}$ of the filter is given by:

$$\text{LoG}^{\text{2D}}_{\sigma_{\text{F},xy}}(x, y) = \frac{x^2 + y^2 - 2\sigma^2_{\text{F},xy}}{\sigma^4_{\text{F},xy}} G^{\text{2D}}_{\sigma_{\text{F},xy}}(x, y). \tag{2.34}$$

Because of the associative property of the convolution operation, applying a LoG filter amounts to applying a Gaussian filter followed by a Laplacian filter. Thus, in principle the LoG filter involves some degree of smoothing of the original image and likewise some enhancement of image regions corresponding to fluorescent particles with a Gaussian-like appearance. The LoG filter also entails evaluating the second derivatives of the image intensities along the image's spatial dimensions. As indicated by (2.34), the kernel has negative values for the relative positions $(x, y) \,|\, x^2 + y^2 < 2\sigma^2_{\text{F},xy}$, i.e., for positions (x, y) inside a circle of radius $\sqrt{2}\sigma_{\text{F},xy}$. Positive values are obtained for positions outside this circle. Over the neighborhood in which the convolution is performed, applying the LoG amounts to computing the difference between the integrated intensity of positions inside the circle and the integrated intensity of positions outside the circle. Image regions

where the difference is large result in a large response for the filter. For a neighborhood corresponding to a single particle, and assuming that the circle of radius $\sqrt{2}\sigma_{\mathrm{F},xy}$ circumscribes the particle, the difference is relatively large since the intensities within the particle differ strongly from that of the background. This results in a large response for the filter. In comparison, image regions corresponding to the background comprise intensity values that are relatively homogeneous, and so the difference is relatively low in comparison to the difference obtained at image regions displaying particles. This also indicates that the standard deviation $\sigma_{\mathrm{F},xy}$ is to be adjusted according to the size of the particles. When applied to fluorescence microscopy images, the LoG filter is known as the spot-enhancing filter (SEF) [160]. Additional approaches for pre-processing the images include non-linear filters (e.g., median filter, anisotropic diffusion [195]), schemes based on mathematical morphology (e.g., the top-hat filter [28], the H-dome transform [179]) as well as schemes based on wavelets (e.g., [138]). Evaluations of several pre-processing schemes for localizing fluorescent particles are presented in [177], [158]. In [177], it turns out that in general no scheme "outperforms all others". The evaluation in [177] also established that schemes based on the multi-scale variance-stabilizing transform (MS-VST) [221], H-dome transform, morphological top-hat, as well as on the SEF are not too sensitive to the corresponding user-defined parameters. This is an important property, since adjusting the parameters for each image is impractical.

In the *particle detection* step, the goal is to determine image regions that correspond to particles. A common technique is to apply a threshold (e.g., [144], [166]) on the intensities of a (pre-processed) image. Thus, pixels with intensity values above the threshold are assumed to represent particles. The result of this technique depends largely on the selected threshold. If the threshold is set too low, false positives might arise; if the threshold is set too high, false negatives might ensue. Determining manually a good threshold value is often impractical and may lead to inconsistent results for different images. Automatic schemes for determining an optimal threshold (e.g., [139]) are therefore required. Usually these schemes compute the threshold as a function of the histogram of the intensity values of an image. For fluorescence microscopy image sequences, the automatic calculation of the threshold should also consider the influence of photobleaching on the distribution of the intensity values.

In the *particle localization* step, the task is to derive an explicit description of each detected particle. Such a description often includes the location of the particle within an image as well as other properties (e.g., mean intensity of the particle). For instance, to compute such a description, one can first identify sets of adjacent pixels (connected components) via a connected-components labeling algorithm (e.g., [78]), which assigns a unique label to each connected component. Each connected component is assumed

to represent one particle and for each particle properties can then be computed. For example, the position can be calculated as the intensity-weighted center-of-mass or the mean intensity can be determined. Also, false positives can be rejected (e.g., [166]) by verifying whether the properties of particles (e.g., pixel count, moments of intensity) fulfill a set of discriminative criteria.

2.2.1.1.2 Top-down Localization Schemes Top-down approaches use model-driven strategies where hypotheses on the possible configuration of the models are tested against the information found in the images. We consider approaches that synthesize an image using a parametric model describing the shape and appearance of a fluorescently labeled particle. A typical model for the appearance of a fluorescence particle is the 2D Gaussian function (e.g. [104], [190], [41], [208]):

$$g_{\text{Gaussian2D}}(x, y) = I_b + (I_{max} - I_b) \exp\left(-\frac{(x - x_p)^2 + (y - y_p)^2}{2\sigma_{xy}^2}\right), \tag{2.35}$$

where I_b represents the background intensity value, I_{max} denotes the peak intensity value, (x_p, y_p) indicates the position of the particle in the image, and σ_{xy} is the standard deviation of the 2D Gaussian function. To model M particles that are in close proximity, a mixture of Gaussian functions may be used [190], [140]:

$$g_{\text{GaussianMix}}(x, y) = I_b + \sum_{m=1}^{M} (I_{max,m} - I_b) \exp\left(-\frac{(x - x_{p,m})^2 + (y - y_{p,m})^2}{2\sigma_{xy,m}^2}\right). \tag{2.36}$$

Here each m-th component is parametrized by its peak intensity value $I_{max,m}$, its position $(x_{p,m}, y_{p,m})$, as well as by its standard deviation $\sigma_{xy,m}$. A 2D Lorentz function (e.g., [52]) has also been used to describe the appearance of a single fluorescence particle:

$$g_{\text{Lorentz2D}}(x, y) = I_b + (I_{max} - I_b) \frac{1}{\left(1 + \frac{(x-x_p)^2 + (y-y_p)^2}{\gamma_{xy}^2}\right)^{\frac{3}{2}}}, \tag{2.37}$$

where γ_{xy} regulates the spread of the function. Analogous to the 2D Gaussian function, a sum of 2D Lorentz functions may be used to describe multiple particles (e.g., [155], [40]). In comparison to the Gaussian, using powers of the Lorentz function also allows adjusting the shape of the function. Numerical evaluation of the function may also be faster. Also, the function is amenable to analytical procedures. Top-down approaches based on such parametric appearance models typically consist of two steps: detection of particle candidates and model fitting. In the first step, typically a search for local intensity maxima is carried out, where each identified maximum represents a candidate

position. Here, the pre-processing schemes described in Section 2.2.1.1.1 may be used. In the second step, for each candidate position the model is fitted to the image intensities. Usually a least-squares fit of the model g_{model} with the observed image intensities g is achieved by minimizing the following objective function:

$$\sum_{(x,y)} (g_{\text{model}}(x,y) - g(x,y))^2, \tag{2.38}$$

where the model g_{model} may be defined by any of the parametric models described above (e.g., the 2D Gaussian function (2.35)). Because of the non-linearity of the models, non-linear minimization algorithms (e.g., Gauss-Newton algorithm, Levenberg-Marquardt algorithm) are required. In images with a low SNR, it has been shown experimentally that fitting a 2D Gaussian model outperforms other bottom-up localization algorithms [36]. Nevertheless, robust bottom-up schemes are required to automatically initialize the candidate positions for the fitting scheme. One issue when fitting a mixture model (e.g., (2.36)) is estimating the correct number of M components, since a large value for M leads to a better description of the image data, yet such a large value may not correspond to the actual number of particles. To address this, the authors in [190] fit an M-component mixture model as well as an $M+1$-component mixture model. For both mixture models, the corresponding residuals and chi-square statistics are computed. A test for significant differences on the chi-square statistics of the models determines whether increasing the number of components is required. In [140], components are iteratively merged and split until the rate of change of the objective function over the iterations is low.

Classification schemes may be also categorized as top-down approaches since the approaches also compare the information found in the images with pre-established information about the particles. For example, in [95], an Adaboost-based classifier using Haar-like features is used to detect fluorescent particles.

2.2.1.2 Motion Correspondence

Deterministic schemes for motion correspondence can be categorized into schemes finding correspondences over two time steps, and schemes finding correspondences over multiple time steps.

One approach [195] for determining correspondences over two time steps involves using a *fuzzy logic* system to determine the degree of correspondence between two measurements. The system considers differences between two measurements, for example, the difference in intensity as well as the difference in position. An *associative rule* of the fuzzy logic system describes the extent of each difference. By using each of these

rules the system determines whether each difference is small or large. To obtain a single value describing the extent of all differences, a *fuzzy centroid* over all the rules is computed and used as a distance function $d_{\text{fuzzy}}(\cdot,\cdot)$ between two measurements. Correspondences are found using a standard nearest neighbor (NN) strategy (see Section 2.1.1.2.1). Other approaches using a nearest neighbor strategy include [6], [171], [107], [144], [105]. In image regions with a high density of particles, the search for correspondences becomes ambiguous, since several possibilities are plausible. In those cases the NN strategy may lead to correspondence errors. To cope with such cases, approaches using a global nearest neighbor (GNN) strategy (cf. Section 2.1.1.2.1) have also been proposed (e.g., [169], [166], [92]). The minimization of the cost function associated with the resulting bi-partite graph is solved for example with algorithms for the transportation problem, the Vogel algorithm, or the Hungarian algorithm. The resulting bi-partite graph may be transformed into a flow network and min-cost max-flow algorithms (e.g., [141]) may be used to find the correspondences (e.g., [198]). Regardless of the scheme used for determining correspondences, approaches using only two consecutive time steps are susceptible to errors of the localization scheme (e.g., missed detections, spurious detections). To cope with detection errors, correspondences over multiple time steps may be established. One approach for finding correspondences over multiple time steps involves constructing a k-partite graph using the measurements obtained over a window comprising k time steps. In this approach correspondences are only allowed between consecutive time steps. The minimization of the cost function C_k is carried out within an integer programming framework [60]. Another approach ([93]) for finding correspondences over multiple time steps obtains an initial set of trajectories over the entire image sequence using a GNN scheme. The approach refines this initial set by extending, merging or splitting trajectories. In this sense such an approach may be viewed as an approach based on tracklets (cf. Section 2.1.3). Note that the initial trajectories (tracklets) themselves are not updated, e.g., erroneous positions within a tracklet are not corrected. Other deterministic approaches taking into account all time steps of an image sequence include [160], [23], [213]. These approaches assume that, since a sequence of 2D images over time may be represented as a 3D spatial-temporal volume $V(x, y, t)$, finding trajectories amounts to finding 3D curves within such a volume. These approaches define an appropriate energy function for such 3D curves using the image data (or products thereof, such as the response of the spot-enhancing filter) as well as prior constraints on the frame-to-frame displacement of the particle. Computing 3D curves that minimize such energy function via dynamic programming algorithms is thus equivalent to finding trajectories. Because the energy is defined for individual 3D curves (trajectories) such approaches can only deal with sparsely positioned particles. In real applications, often

the particles are densely positioned and this limits the applicability of these type of approaches.

In summary, deterministic approaches ignore the uncertainty involved in the behavior of the fluorescent particles as well as in the imaging system; their applicability thus remains limited to relatively straightforward scenarios (e.g., low object density, high SNR).

2.2.2 Probabilistic Approaches

Formulated within a Bayesian framework (cf. Section 2.1.2.1), probabilistic approaches cope with the uncertainty by defining a posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ on the variables $\mathbf{x}_t$ describing the fluorescent particles given a series of image-derived measurements $\mathbf{y}_{1:t}$. This posterior can be resolved via a Kalman filter or a particle filter. Approaches based on the Kalman filter (e.g., [64], [38], [98], [123], [215]) incorporate only the measurements computed by a bottom-up localization algorithm (cf. Section 2.2.1.1.1); thus localization and position estimation are uncoupled steps. In contrast, the particle filter (e.g., [176], [218], [70], [32], [75]) queries directly multiple image positions (represented by random samples) to determine the location of an object. The filter thus combines localization and position estimation and achieves more robust results (e.g., [70]). Rao-Blackwellized particle filters have also been used (e.g., [178]) to track fluorescent particles. Because of the random nature of its top-down localization scheme, the particle filter uses a relatively large number of samples. This introduces a high computational cost because an image likelihood is evaluated for each sample. While one can improve the efficiency of the particle filter, (e.g., by exploring hierarchically the feature space [214], by biasing the samples towards regions with high likelihoods [30], or by marginalization [99]), several hundred samples are still required to ensure convergence. An approach based on Kalman filters combining a bottom-up localization scheme with a top-down localization strategy has been presented in [179], but there the combined localization scheme does not exploit the appearance parameters of each object. Probabilistic approaches have also been proposed for addressing the uncertainty involved in the task of motion correspondence [42]. These include approaches based on the joint probabilistic data association filter (JPDAF, e.g., [179], [98], [154]), the multiple hypothesis tracking algorithm (MHT, e.g., [38], [115]) as well as approaches sampling correspondences via Monte Carlo schemes (e.g., [74]). These approaches explore exhaustively the space of correspondences over two or multiple time steps, which entails a large computational overhead when tracking a very large number of objects.

2.3 Approaches for Behavior Identification in Computer Vision

Given an image sequence, the task of behavior identification entails assigning a *behavior label* to representations extracted from the image data [145]. Thus, approaches for behavior identification often deal with *representation* and *representation extraction* as well as *label computation* [33]. In this context, a behavior label denotes a symbol carrying a semantic connotation related to a certain action or behavior. The extracted representations may correspond to low-level features (e.g., feature points ([205], [110], [53]), optical flow vectors (e.g., [58]), and spatial-temporal volumes (e.g., [20]). Approaches using such representations detect behaviors at the *image level*. Alternatively, the representations may correspond to trajectories describing the motion of objects in an image sequence (e.g., [175], [76], [37], [194]). In this case the behavior is detected at the *object level*. Because the representations are extracted from image sequences, the representations also encode temporal information. The temporal information may have a *global*, *semi-global*, or *local* scope, and approaches are designed to deal with a particular temporal scope. In computer vision, several approaches for behavior identification have been developed, especially for identifying behaviors related to humans. This is motivated by applications of such approaches in the fields of surveillance (e.g., [186], [46], [131], [130]), gait analysis (e.g., [26], [201], [17], [48]), traffic analysis (e.g., [85], [205], [161], [83]), sports (e.g., [58], [118]), face expression analysis (e.g., [54], [106]), hand gesture analysis (e.g., [148], [13], [76]) as well as general identification of human actions (e.g., drinking coffee as in [111]). There is a large number of approaches for the identification of human behaviors (see for example the recent surveys in [206], [31], [145], [127], [84]). Animals are also the subjects of approaches for behavior identification. Approaches for the identification of behaviors of bees (e.g., [134], [200]) as well as rodents (e.g., [53]) have been proposed.

In this section we review approaches for behavior identification. We focus on the task of label computation for object-based representations, since these are readily extracted by a tracking approach. We concentrate on algorithms grounded within a Bayesian framework.

2.3.1 Label Computation

We focus on approaches computing labels for object-based representations with local scopes. A Bayesian approach for computing labels involves finding the most probable sequence of labels given the observed local representations. Hidden Markov models (HMMs) (e.g., [210]) are typically used for finding such sequences of labels. HMMs spec-

ify transition probabilities over the labels and thus only plausible behavior sequences are evaluated. One disadvantage of HMMs is that the hidden variables (i.e., the behaviors) may only take discrete values. Moreover, HMMs describe the observed temporal representations using piecewise constant functions. Approaches based on *stochastic hybrid systems* overcome these limitations by estimating both the discrete behavior as well the continuous representations.

Stochastic hybrid systems have first been introduced in automatic control applications (e.g., [193]), and have been recently revisited for behavior identification in the field of computer vision. Applications include identification of behaviors based on face expressions (e.g., [54], [106]), as well as on human gaits and pose (e.g., [17], [48]). Other applications include the identification of the behavior of bees (e.g., [134], [200]). In biological imaging, hybrid stochastic systems have been used for estimating the position of fluorescent particles (e.g., [64], [178]) or cells (e.g., [113]), however, in these approaches the hybrid stochastic system does not exploit the intensity information of the biological objects and the estimated discrete variables are not used for behavior identification. In general, estimation of the hybrid states is often carried out within a Bayesian framework where the aim is to compute a posterior distribution on the hybrid states given the observed temporal representations of the object under consideration. This task is challenging since the optimal solution entails an exponential computational effort with respect to time. Therefore, approximate algorithms have been proposed. For example, [200] suggested a greedy approach that involves first approximating the posterior distribution of the continuous variables followed by calculating maximum likelihood (ML) estimates for the discrete behavior variables. While this approach significantly reduces the complexity of the estimation task, the greedy ML estimates do not consider the validity of transitions between behaviors. Another approach uses an approximation of the posterior via a mixture of Gaussians (e.g., [17]). However, this approach is only applicable to linear Gaussian models. In [134], a data-driven Markov Chain Monte Carlo (MCMC) method is used to approximate the posterior. There, the MCMC proposal distribution is derived from the likelihood of the observed temporal representations as well as from the transition probabilities between different behaviors. While the solution space is efficiently explored, the approach requires a relatively large amount of training data that should include a label for each behavior at each time step. Additionally, MCMC methods do not typically support sequential analysis of the observed temporal representations, which limits their applicability to off-line inference tasks.

Particle filters (e.g., [88]) provide a sequential alternative to MCMC methods. Within computer vision, the application of particle filters for carrying out inference in stochastic hybrid systems is first described in [87]. Because of its applicability to non-linear

and non-Gaussian models as well as its recognition-by-synthesis approach, such a *hybrid particle filter* has seen increasing interest for identifying behaviors. For example, [18] represent behaviors (e.g., facial expressions) as temporal trajectories within a certain space (e.g., a space defined by the basis of optical flow fields). The temporal trajectories are sequentially matched to a temporal neighborhood of the observed temporal representations via a hybrid particle filter, where the temporal neighborhood is defined by a temporal window. However, the length of this window influences the results and determining the optimal length is not trivial. In certain applications, the transition probabilities for certain behaviors may be low, which entails that unlikely behaviors may be supported by very few hybrid samples. The lack of support in certain regions of the discrete space reduces the accuracy of the approximated hybrid posterior. One strategy to cope with this issue involves carrying out importance sampling on the discrete variable only, whereby the discrete space is explored more effectively by using an importance transition matrix ([156]). However, performing importance sampling on the discrete variable only may generally decrease the accuracy of the estimates for the continuous variables. Another strategy to cope with the lack of support for unlikely behaviors is to increase the number of samples. This straightforward strategy works well at the expense of an increase in the computational cost, which scales linearly with respect to the number of samples. To compensate for the high computational cost, some approaches (e.g., [226]) assume that the behavior remains fixed over time. Such an approach reduces significantly the extent of the solution space but it is not applicable to the task of identifying disparate behaviors over time. Other schemes (e.g., [106]) assume no direct dependence between the discrete variable and the continuous variables, but such an assumption may not hold in many applications.

An additional issue with the hybrid particle filter is that the estimates for the variables are computed sequentially using only the temporal representations found up to the current time step. Using all available temporal representations may increase the accuracy of the computed estimates. A *smoothing* algorithm based on particle filters for computing discrete and continuous variables based on all available temporal representations has been presented in [133]. However, the computational cost of such a smoothing scheme is quadratic with respect to the number of samples. Most schemes based on the hybrid particle filter deal with one object only, so smoothing may be applicable in those cases. However, in cases where multiple objects are present over several time steps, such a smoothing scheme may be impractical.

2.4 Approaches for Behavior Identification of Fluorescent Particles

In this section, we concentrate on previous work identifying behaviors based on the appearance of the fluorescent particles. We also review approaches for identifying fusion of fluorescent particles.

Behaviors of fluorescent particles, such as fusion, are often reflected as fluctuations of the temporal intensity statistics of individual particles. Changes in the intensity over time may be detected at the image level by computing the difference image between two consecutive images using single pixels (e.g., [132], [170], [125]) or image regions (e.g., [24]). Since changes in the intensity may also arise from other phenomena (e.g., motion of the particles, image noise), approaches based on image differencing may not discriminate accurately between changes in the intensity arising from behaviors of interest (e.g., fusion) and changes originating from other phenomena. Also, these approaches are typically not applicable for detecting behaviors that entail changes in temporal statistics other than the intensity (e.g., size of a particle). By tracking fluorescent particles and thereby obtaining the temporal statistics of each object, the changes can be detected at the object level. For example, within the context of neurobiology, the fusion of a fluorescent neurotransmitter vesicle with the cell membrane is detected by fitting a linear model to the curve of estimated radii over time of the fluorescent particle ([41]). A statistical test on the fitted slope parameter, which reflects the rate of change in the particle's radii, determines whether fusion occurred. Such a *global* approach, while taking advantage of all temporal data, assumes a constant behavior and is not able to identify potentially disparate behaviors over time. In comparison, *local* approaches identify behaviors of interest at each time step by taking into account the statistics of few consecutive time steps.

Local approaches that detect fusion based on the intensity over time of an individual particle typically exploit the fact that this behavior entails a rapid change in the intensity. For example, [61] used a derivative-based approach, where fusion of an influenza virus particle with the cell membrane is represented by large values of the first derivative of the particle's intensity with respect to time. To numerically estimate the derivatives typically small temporal neighborhoods are incorporated and therefore these approaches are susceptible to noise. In [11], [51] the rate of change in intensity has been used for detecting fusion of glucose transporter 4 (GLUT4) vesicles with the cell membrane. Local approaches do not take into account results from previous time steps for analyzing subsequent time steps, i.e., the temporal coherence of the particle's behavior is not exploited. Approaches based on Hidden Markov Models (HMMs) address this

shortcoming by specifying local transition probabilities between different behaviors.

2.5 Summary

In this chapter, we have reviewed previous work for tracking multiple particles as well as for identifying behaviors of fluorescent particles. Tracking approaches are classified into deterministic or probabilistic schemes. Behavior identification schemes are classified based on their temporal scope and also based on the underlying approach for label calculation. For tracking, probabilistic approaches deliver more robust results than deterministic approaches. We delve deeper into probabilistic tracking approaches in the next chapter. For behavior identification, we focus on hybrid stochastic systems, since these estimate both discrete and continuous variables. Our approach for behavior identification is presented in Chapter 4.

Chapter 3

Tracking Fluorescent Particles

In this chapter, we describe our approaches for tracking multiple fluorescent particles. We first discuss the tasks of object localization and motion correspondence. Then we delve into probabilistic approaches grounded within the theory of Bayesian sequential estimation. We introduce approaches based on the Kalman filter, and independent particle filters [70]. We also present our approach based on Probabilistic Data Association (PDA) [68].

3.1 Spot Detection and Localization via the Spot-Enhancing Filter

In fluorescence microscopy images, the intensity structure corresponding to a fluorescent particle typically resembles a 2D Gaussian function, where the peak intensity of the particle is larger than the intensity of the background. One approach for localizing such particles is threshold-based segmentation. However, due to image noise and other biological structures with similar intensities (e.g., cellular autofluorescence), this approach generally leads to a high number of false detections. To improve object localization, we use an algorithm based on the spot-enhancing filter (SEF) ([160]). The algorithm, which consists in convolving an image with a Laplacian-of-Gaussian (LoG) kernel that is parametrized by the standard deviation $\sigma_{\mathrm{LoG},xy}$ (cf. (2.34) as well as Figure 3.1), has the property that it enhances the intensities of Gaussian-like particles while simultaneously suppressing noise, as well as attenuating the intensities of background structures. After filtering, we apply an intensity threshold k_{intens}, which is computed by

$$k_{\mathrm{intens}} = \mu_{\mathrm{intens}} + c\,\sigma_{\mathrm{intens}}, \tag{3.1}$$

where μ_{intens} is the mean intensity of the image, σ_{intens} the standard deviation, and c a user-defined factor. In our case, we compute a threshold value for each image of an image sequence, since the intensity distribution generally changes over time. To identify fluorescent particles, we employ a connected components labeling algorithm ([78]) assuming 8-connectivity. False positives are rejected by enforcing a minimum and maximum number of pixels for each detected particle.

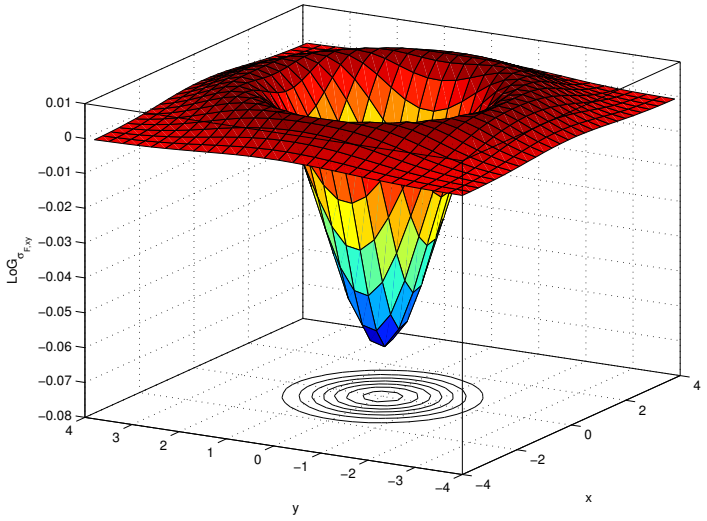


Figure 3.1: Kernel of the Laplacian-of-Gaussian (LoG) filter ($\sigma_{\text{F},xy} = 1.5$ pixels).

3.2 Spot Detection and Localization via 2D Gaussian Fitting

Under the assumption that the intensity distribution of a fluorescent particle can be represented by a 2D Gaussian function (see (2.35)), a natural approach for particle localization consists in fitting such a function to candidate regions-of-interest (ROIs) determined by local intensity maxima (e.g., [190], [208]) (see Figure 3.2 for a diagrammatic representation). Prior to local maxima finding, we use intensity clipping and a Gaussian filter for noise reduction. For each candidate ROI we apply 2D Gaussian fitting via a least-squares estimator. The final step consists in rejecting false positives based on the following set of criteria: maximum ellipticity of the fitted function, minimum total intensity, minimum contrast, and maximum pixel count.

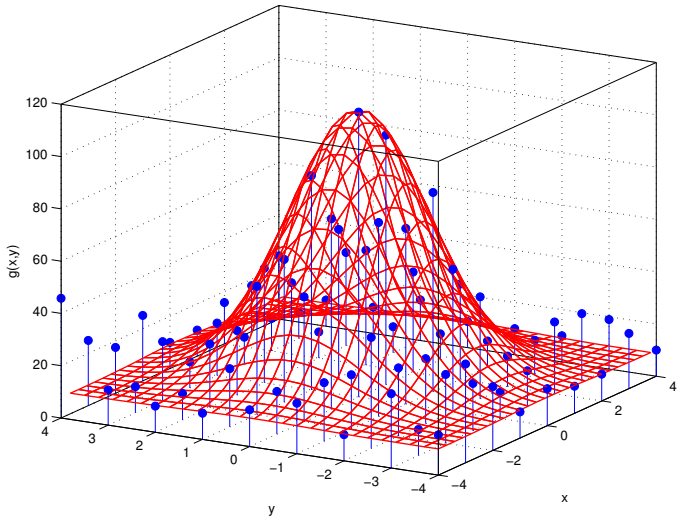


Figure 3.2: Fitting a 2D Gaussian function (red) to image intensity values (blue).

3.3 Motion Correspondence via Global Nearest Neighbor

When applied to an image displaying multiple fluorescent particles, a spot detection scheme yields a set $Y_{\mathrm{BU}} \triangleq \{\mathbf{y}_{\mathrm{BU},1}, \mathbf{y}_{\mathrm{BU},2}, \ldots, \mathbf{y}_{\mathrm{BU},N_{m,\mathrm{BU}}}\}$ of $N_{m,\mathrm{BU}}$ bottom-up (BU) measurements. To assign a single measurement to each tracked object, we find one-to-one correspondences between the set of bottom-up measurements Y_{BU} and the set $\hat{Y} \triangleq \{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_{N_{\mathrm{obj}}}\}$ of N_{obj} predicted measurements. To solve this correspondence problem, we use a *global* nearest neighbor (GNN) approach [166] based on a graph-theoretical approach for the transportation problem (see Section 2.1.1.2.1). The algorithm encodes the correspondences via an *association matrix* $\mathbf{A}$ (of dimensions $N_{\mathrm{obj}} \times N_{m,\mathrm{BU}}$) where the rows correspond to the predictions and the columns represent the measurements. The entries a_{ij} of the matrix $\mathbf{A}$ denote whether a match exists between the prediction $\hat{\mathbf{y}}_i$ and the measurement $\mathbf{y}_{\mathrm{BU},j}$. In our case, a_{ij} is set to '1' if the prediction $\hat{\mathbf{y}}_i$ and measurement $\mathbf{y}_{\mathrm{BU},j}$ correspond to each other; otherwise a_{ij} is set to '0':

$$a_{ij} = \begin{cases} 1 & \text{if } \hat{\mathbf{y}}_i \text{ and } \mathbf{y}_{\mathrm{BU},j} \text{ match} \\ 0 & \text{otherwise} \end{cases} . \tag{3.2}$$

Since we restrict ourselves to one-to-one assignments, the association matrix $\mathbf{A}$ is subject to the topological constraint that each row and column may have at most one '1'. Also, since generally the number of predictions may not match the number of measurements, i.e., $N_{\mathrm{obj}} \neq N_{m,\mathrm{BU}}$, it follows that some of these may remain unmatched. To solve the

practical difficulties that this inequality introduces, the notions of a dummy prediction $\hat{\mathbf{y}}_0$ and a dummy measurement $\mathbf{y}_{\mathrm{BU},0}$ are introduced. The purpose of these dummies is the following: if a prediction $\hat{\mathbf{y}}_i$ remains unmatched, it will be assigned to the dummy measurement $\mathbf{y}_{\mathrm{BU},0}$. Analogously, if a measurement $\mathbf{y}_{\mathrm{BU},j}$ remains unmatched, it will be assigned to the dummy prediction $\hat{\mathbf{y}}_0$. The introduction of the dummies entails the augmentation of the matrix $\mathbf{A}$ by one row and one column. For this particular row and column, the above-mentioned topological constraint does not hold, since, for instance, several measurements may remain unmatched.

A global nearest neighbor algorithm aims at finding the correspondences between both sets by adjusting the association matrix $\mathbf{A}$ such that the total displacement d_{total} induced by the set of correspondences is minimized. Let $d(\hat{\mathbf{y}}_i, \mathbf{y}_{\mathrm{BU},j})$ denote the Euclidean distance in $\mathbb{R}^2$. The total displacement d_{total} is defined as follows:

$$d_{\mathrm{total}} = \sum_{i=0}^{N_{\mathrm{obj}}} \sum_{j=0}^{N_{m,\mathrm{BU}}} a_{ij}\, d(\hat{\mathbf{y}}_i, \mathbf{y}_{\mathrm{BU},j}). \tag{3.3}$$

To reduce the complexity of the task, only the most plausible matches (in terms of the distance $d(\cdot,\cdot)$) are considered. This is done by disregarding matches that induce a relatively large distance. To this end, we discard those matches that introduce a displacement larger than the expected maximum displacement d_{max} ([199]). Additionally, the distance between a prediction $\hat{\mathbf{y}}_i$ and a dummy measurement $\mathbf{y}_{\mathrm{BU},0}$ as well as the distance between a dummy prediction $\hat{\mathbf{y}}_0$ and a measurement $\mathbf{y}_{\mathrm{BU},j}$ is set equal to d_{max}.

The motion correspondence algorithm proceeds in two stages: first, initial assignments are created via a greedy nearest neighbor approach. In the second stage, the optimal set of assignments (optimal in the sense that yields a minimal d_{total}). To achieve the latter, we employ the optimization algorithm presented in [166], which is based on a graph-theoretical approach for the transportation problem.

For tracking, one needs to determine the start and end points of each trajectory. The dummies, in principle, provide the means through which such start and end points can be determined. Intuitively, a measurement assigned to a dummy prediction $\hat{\mathbf{y}}_0$ denotes the start of a new trajectory. Similarly, a prediction matched to a dummy measurement $\mathbf{y}_{\mathrm{BU},0}$ denotes the end of a trajectory.

3.4 Bayesian Framework

In our work, tracking is formulated as a Bayesian sequential estimation problem. We recall here the theory within the context of tracking fluorescent particles. Within a *one-body state space*, it is assumed that a fluorescent particle is represented by a state vector

$\mathbf{x}_t$ and that a noisy measurement $\mathbf{y}_t$ reflects the true state of $\mathbf{x}_t$. At time step t, the aim is to estimate the state $\mathbf{x}_t$ of a fluorescent particle given a sequence of measurements $\mathbf{y}_{1:t}$. By modeling the temporal evolution using a *dynamical model* $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and incorporating measurements derived from the images via a *measurement model* $p(\mathbf{y}_t|\mathbf{x}_t)$, a Bayesian filter estimates the *posterior* distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ via stochastic propagation and Bayes' theorem:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, d\mathbf{x}_{t-1} \tag{3.4}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{y}_{1:t-1}). \tag{3.5}$$

An estimate of $\mathbf{x}_t$ can be obtained from the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. For linear and Gaussian models, one can resolve analytically the posterior using a Kalman filter; for non-linear and/or non-parametric models, the particle filter provides a numerical solution.

3.5 Kalman Filter

The Kalman filter (e.g., [97],[5]) represents the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ via a Gaussian probability distribution that is parametrized by its mean vector $\mathbf{m}_t$ and its covariance matrix $\mathbf{P}_t$:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\cdot; \mathbf{m}_t, \mathbf{P}_t), \tag{3.6}$$

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The Kalman filter also assumes that the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ as well as the measurement model $p(\mathbf{y}_t|\mathbf{x}_t)$ are linear and Gaussian:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{F}\mathbf{x}_{t-1}, \mathbf{Q}) \tag{3.7}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t; \mathbf{H}\mathbf{x}_t, \mathbf{R}), \tag{3.8}$$

where the *transition* matrix $\mathbf{F}$ and the *measurement* matrix $\mathbf{H}$ are known matrices. The covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ encode the uncertainty about the prediction and the measurement, respectively. Given a posterior $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ at time step $t-1$, the posterior at time step t can be computed by first carrying out a prediction on the state vector as well as on the associated covariance matrix:

$$\hat{\mathbf{m}} = \mathbf{F}\mathbf{m}_{t-1} \tag{3.9}$$

$$\hat{\mathbf{P}} = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \mathbf{Q}. \tag{3.10}$$

To improve the readability we omit the time step t index in the following. The predicted state vector $\hat{\mathbf{m}}$ can be transformed onto the measurement space to obtain a predicted measurement $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = \mathbf{H}\hat{\mathbf{m}}. \tag{3.11}$$

The predicted errors as encoded by the predicted covariance matrix $\hat{\mathbf{P}}$ can also be propagated onto the measurement space to obtain a predicted measurement covariance $\mathbf{S}$, which also takes into account the measurement noise process as encoded by $\mathbf{R}$:

$$\mathbf{S} = \mathbf{H}\hat{\mathbf{P}}\mathbf{H}^T + \mathbf{R}. \tag{3.12}$$

Once a measurement $\mathbf{y}$ is derived from the image at time step t, the difference between the actual measurement and the predicted measurement, i.e., the *innovation* $\boldsymbol{\nu}$, is calculated:

$$\boldsymbol{\nu} = \mathbf{y} - \hat{\mathbf{y}}. \tag{3.13}$$

The Kalman gain $\mathbf{K}$ is computed as follows:

$$\mathbf{K} = \hat{\mathbf{P}}\mathbf{H}^T\mathbf{S}^{-1}. \tag{3.14}$$

An estimate for the mean vector $\mathbf{m}$ is finally calculated by correcting the prediction $\hat{\mathbf{m}}$ with the innovation $\boldsymbol{\nu}$, where the correction is regulated by the Kalman gain $\mathbf{K}$:

$$\mathbf{m} = \hat{\mathbf{m}} + \mathbf{K}\boldsymbol{\nu}. \tag{3.15}$$

The covariance matrix $\mathbf{P}$ can also be calculated by updating the predicted covariance matrix $\hat{\mathbf{P}}$ with the Kalman gain $\mathbf{K}$:

$$\mathbf{P} = (\mathbf{I} - \mathbf{K}\mathbf{H})\hat{\mathbf{P}}, \tag{3.16}$$

where $\mathbf{I}$ is the identity matrix.

3.6 Particle Filter

The particle filter (e.g., [72], [88]) is a numerical approach that has been quite successful for non-linear and/or non-Gaussian models within the context of visual tracking, e.g., human tracking ([220], [109], [151]), face tracking ([227]), sports tracking ([203], [136], [30]), animal tracking ([196], [100]), as well as vision-based simultaneous localization and mapping for autonomous robots ([191]). In our application, fluorescent particles move in a relatively unpredictable manner, and therefore the particle filter is well-suited.

The key idea behind this approach is that the posterior probability density function (pdf) $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is approximated with a set of random samples (the 'particles') that are associated with importance weights. Let $\{\mathbf{x}_t^i; w_t^i\}_{i=1}^{N_s}$ represent such a set, where N_s denotes its size, $\mathbf{x}_t^i$ is a sample and w_t^i represents the corresponding (normalized) weight. Using this particle set, the posterior distribution can be approximated as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N_s} w_t^i \, \delta(\mathbf{x}_t - \mathbf{x}_t^i), \tag{3.17}$$

where $\delta(\cdot)$ denotes the Dirac delta measure and $\sum_{i=1}^{N_s} w_t^i = 1$. The particle filter accomplishes the Bayesian recursion in (3.4) and (3.5) in a two-step fashion. First, candidate samples $\{\mathbf{x}_t^i\}_{i=1}^{N_s}$ are generated from a *proposal distribution* $q(\cdot)$, which is typically set to be equivalent to the dynamical model, i.e.,

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}). \tag{3.18}$$

That is, the samples are propagated over time by carrying out predictions by means of the dynamical model. More convenient proposal distributions, i.e., proposal distributions that explore more efficiently the state space, may be employed; we delve further into this point below. The second step of the algorithm involves reweighing the samples. The weight of each sample $\mathbf{x}_t^i$ is given by the following importance ratio:

$$w_t^i \propto w_{t-1}^i \frac{p(\mathbf{y}_t|\mathbf{x}_t^i)\, p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}. \tag{3.19}$$

The importance ratio not only measures the plausibility of the predicted sample $\mathbf{x}_t^i$ by comparing it to the obtained measurement $\mathbf{y}_t$, but also compensates for any bias introduced by the proposal distribution $q(\cdot)$. At certain time steps, a resampling step may be required to maintain only the most likely samples (hypotheses), thereby avoiding the *degeneracy problem* ([55], [9]). At each time step, an estimate of the object's state may be derived by, for instance, computing the mean of the approximated posterior pdf:

$$\bar{\mathbf{x}}_t = \sum_{i=1}^{N_s} w_t^i \, \mathbf{x}_t^i. \tag{3.20}$$

3.6.1 Boosted Particle Filter

The performance of the particle filter may be enhanced by selecting an appropriate proposal distribution $q(\cdot)$. For visual tracking, the proposal distribution can take into account image information to explore regions of the state space with high likelihoods. For instance, candidate samples may be generated around regions where objects have been

located via a (deterministic) object localization algorithm. A particle filter employing such a proposal distribution is denoted as a *boosted particle filter* ([136]).

More concretely, the proposal distribution $q(\cdot)$ of the boosted particle filter is defined by taking into account both the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and the 'boosted' distribution $q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t})$ resulting from the measurement $\mathbf{y}_{\text{BU},t}$ obtained by a bottom-up (BU) particle localization algorithm (e.g., the approach based on the spot-enhancing filter in Section 3.1). The boosted distribution $q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t})$ is generated by superimposing a Gaussian distribution centered at $\mathbf{y}_{\text{BU},t}$ with a diagonal covariance matrix $\mathbf{Q}_{\text{boost}}$:

$$q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t}) = \mathcal{N}(\mathbf{x}_t; \mathbf{y}_{\text{BU}}, \mathbf{Q}_{\text{boost}}), \tag{3.21}$$

where $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The mixture proposal distribution $q_{\text{mix}}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{\text{BU},1:t})$ is now defined as the mixture between the boosted distribution $q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t})$ and the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$:

$$q_{\text{mix}}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{\text{BU},1:t}) = \alpha_{\text{boost}}\, q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t}) + (1 - \alpha_{\text{boost}})\, p(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{3.22}$$

where the parameter α_{boost} regulates the influence of the boosted distribution $q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t})$. If $\alpha_{\text{boost}} = 0$, the proposal distribution reduces to the traditional proposal distribution of the particle filter. The bias introduced by this proposal distribution must be compensated in the calculation of the importance weights compared to (3.19):

$$w_t^i \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^i)\, p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{q_{\text{mix}}(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{y}_{\text{BU},1:t})} \tag{3.23}$$

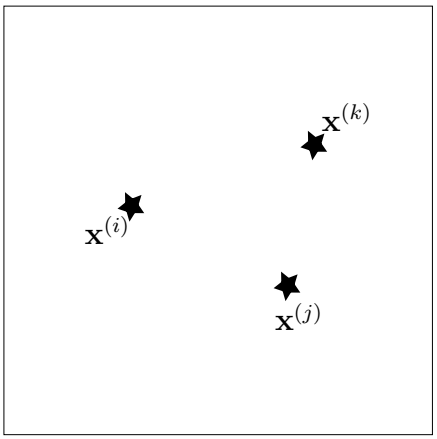
$$\propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^i) p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{\alpha_{\text{boost}}\, q_{\text{boost}}(\mathbf{x}_t^i|\mathbf{y}_{\text{BU},t}) + (1 - \alpha_{\text{boost}})\, p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}. \tag{3.24}$$

One must note that the mixture proposal distribution $q_{\text{mix}}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{\text{BU},1:t})$ is only employed if the boosted distribution $q_{\text{boost}}(\mathbf{x}_t|\mathbf{y}_{\text{BU},t})$ and the distribution of the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ overlap. If there is no overlap, then α_{boost} is set to zero, so that only the distribution generated by the dynamical model is taken into account. In general, the value for the parameter α_{boost} is determined experimentally. The boosted particle filter thus provides the means to explicitly combine a particle filter with measured position estimates obtained by a bottom-up localization algorithm.

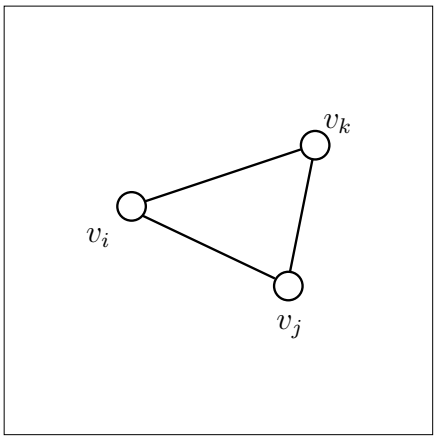
3.6.2 Tracking Multiple Objects with Independent Particle Filters

One approach for tracking multiple objects that have a similar appearance consists in instantiating one particle filter per object. This approach is denoted as *independent particle filters* (IPF) [71]. The IPF operates on a relatively small state space, thereby attaining a good approximation of the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ with relatively few samples. This entails low computational demands, which is a key property for our application, where often a relatively large number of objects needs to be tracked simultaneously. Note that the accuracy of the estimate of $\mathbf{x}_t$ does not deteriorate as the number of objects increases, since each filter instantiates an independent set of particles of size N_s. The approach based on IPF works well if the distance between tracked objects is relatively large. However, the approach generally fails in cases where objects pass close to each other, since the corresponding filters, being unaware of each other, converge towards the object with the best score in terms of $p(\mathbf{y}_t|\mathbf{x}_t)$, i.e., the best likelihood. Approaches to address this problem have been proposed by [146] and [30]. In [146], an exclusion mechanism based on a magnetic potential model has been proposed, and the approach has been applied for tracking human faces. In [30], the mean-shift algorithm has been used for deterministically biasing the particles; the approach has been used for tracking hockey players. The former approach prevents objects from merging, which is not desirable in our application, while the mean-shift approach might yield an incorrect offset due to the close proximity of multiple objects with a similar appearance. In our work, we developed a *penalization scheme* that is based on both probabilistic and deterministic information, and which does not necessarily preclude objects from merging (see Figure 3.3 for a schematic representation). Our penalization scheme comprises three steps: first, the scheme determines objects that are in close proximity. This reduces to finding cliques in an undirected graph $\Gamma = (V, E)$, where a vertex v_i is defined by the filtered position estimate of object i, and an edge $e = \{v_i, v_j\}$ is said to join vertices v_i and v_j if the Euclidean distance between the positions of the two objects is below a predefined value. The second step determines the most plausible position $\hat{\mathbf{x}}_t$ for each object in each clique. For this purpose, modes are sought in the probability density function that is induced by merging all particles of all filters of one clique: given k objects, such a distribution exhibits k modes. In some cases, k modes might not be obtainable, for instance, if objects are too close to one another. The assignment of modes to objects is carried out via a deterministic motion correspondence algorithm, namely a global nearest neighbor approach [166]. If no mode (i.e., no most plausible position) is assigned to an object, this object is not further considered in the penalization scheme. This entails that the penalization scheme may allow filters to coalesce for a while. In

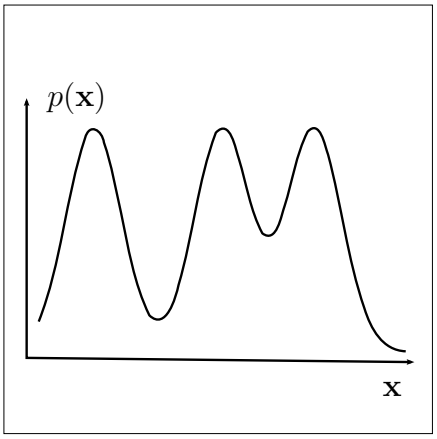
the third step, the weights of those misleading particles that are relatively distant to the most plausible position $\hat{\mathbf{x}}_t$ of an object are assigned lower values via a Gaussian function centered at $\hat{\mathbf{x}}_t$ with a standard deviation $\sigma_{penalize}$. Given the lower weights, the resampling step of the particle filter may discard the misleading particles, thereby preventing filters from coalescing.



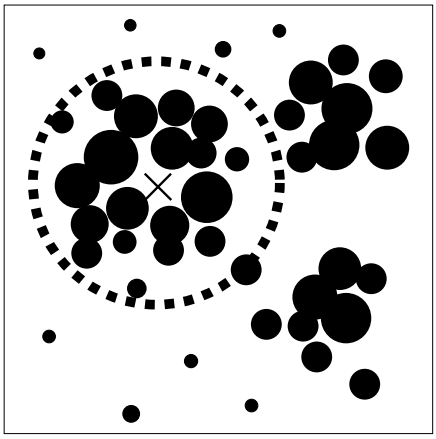
(a) Filtered estimates for three objects.



(b) Graph Γ derived from the filtered estimates. Cliques in this graph are assumed to represent objects in close proximity.



(c) Representation of the probability distribution that is induced by merging all samples from a clique. Modes in this distribution are assumed to represent the most plausible positions for the objects involved in a clique.



(d) Particles representing the filtered posterior for object i: each particle is represented by a solid circle and the corresponding weight is proportional to the circle's radius. The assigned mode $\hat{\mathbf{x}}$ is shown as a cross and the ball B_r is shown as a dashed circle. Particles outside the ball are subsequently penalized using a Gaussian scheme.

Figure 3.3: Diagram illustrating the different steps of the penalization scheme.

3.7 Probabilistic Data Association

The particle filter evaluates an image likelihood for each of its random samples. Because of numerical issues, the particle filters requires several hundred samples. Thus several hundred evaluations of the image likelihood are performed which is computationally expensive. This precludes the efficient application of the particle filter in real applications where the number of objects is very high (e.g., 500 objects). Here we introduce an efficient and robust approach for tracking multiple fluorescent particles based on *probabilistic data association* (PDA). We propose a localization scheme that combines bottom-up and top-down strategies. The bottom-up strategy relies on a spot detection scheme (e.g., the spot-enhancing filter [160]) while the top-down strategy generates measurements via an ellipsoidal sampling scheme based on the Gaussian probabilities calculated by the Kalman filter. Since we use an elliptical sampling scheme we refer to our approach as *PDAE*. The localization scheme generates multiple measurements which are integrated into the Kalman filter using the principle of a combined innovation as proposed in the PDA algorithm (e.g., [102], [149]). Unlike the standard PDA algorithm, our PDAE approach interprets the association probabilities of each measurement as weights relative to the image likelihood of the object. To calculate the weights, we synthesize hypotheses on the possible appearance and location of the object and test the hypotheses against the information found in the images. Thus, rather than using the combined innovation for addressing the correspondence problem, we use the combined innovation as a localization mechanism based on a *recognition-by-synthesis* scheme. Since the image likelihood considers only one object, an image region may support the measurements of multiple neighboring objects. Thus, multiple filters may converge towards the same image regions. To discourage this, we calculate the support of each image position relative to the neighboring objects of a tracked object and use this relative support to re-calculate the probabilities of each measurement of the tracked object. To incorporate multiple motion models, we combine our PDAE approach with the interacting multiple model (IMM) algorithm (e.g., [21], [64]). Analogous to the particle filter (e.g., [176]), our approach capitalizes on multiple measurements, uses the image data directly, and performs both localization and position estimation over time. Unlike previous linear schemes (e.g., [162], [101] for microtubuli) or radial schemes (e.g., [44] for white blood cells, or [153] for *C. elegans* worms) for localization in biological imaging, our ellipsoidal sampling scheme for generating measurements takes into consideration the directions along which the main deviations from the predicted position are expected. Also, the size of the ellipsoid is automatically determined. Moreover, our approach can be used for 2D and 3D images. In contrast to previous approaches using the standard PDA algorithm, our approach does not rely exclusively on bottom-up

measurements (e.g., [98], [154]) nor does it rely on random sampling (e.g., [179], [149]). In contrast to approaches based on multiple hypothesis tracking (e.g., [38], [115]), the computational cost when dealing with multiple objects is at most quadratic relative to the number of objects (compared to exponential costs). In comparison to approaches based on the IMM algorithm (e.g., [64], [60], [154]), our approach uses a top-down localization strategy, and to compute the likelihood of the motion models, our approach does not assume Gaussian densities. To the best of our knowledge, this is the first time that a tracking approach based on the PDA algorithm using a recognition-by-synthesis scheme underpinned by an ellipsoidal sampling scheme that generates bottom-up as well as top-down measurements is introduced in conjunction with either the Kalman filter or the IMM algorithm to track multiple fluorescent particles in 2D and 3D microscopy image sequences.

3.7.1 Measurement Process

3.7.1.1 Overview

In biological imaging, tracking approaches based on the Kalman filter typically use only a single measurement $\mathbf{y}$ in (3.13) to update each filter (e.g., [64], [113], [38], [123]). In contrast, we consider *multiple* measurements (e.g., [98], [179]) to update each Kalman filter. In our approach, these measurements are obtained by considering the image data (i.e., bottom-up localization) as well as by using a localization scheme that takes into account the predicted measurement $\hat{\mathbf{y}}$ (i.e., top-down localization). The measurements are assimilated by the Kalman filter via the combined innovation principle of the probabilistic data association (PDA) algorithm. Below, we describe our schemes for bottom-up and top-down localization as well as the scheme for measurement integration via the PDA algorithm. In the following, we describe the measurement process considering only a single particle. The case of multiple particles is considered in Section 3.7.2 below.

3.7.1.2 Bottom-up Localization

For detecting and localizing a particle based on the image data only, any spot detection scheme may be used (see [177], [158] for an evaluation of spot detection schemes in fluorescence microscopy). For tracking virus particles, we use the approach based on the spot-enhancing filter outline above (see Section 3.1). For elongated objects we use the approach based on Gaussian fitting (see Section 3.2). Applying the spot detection scheme to the image ideally yields a single bottom-up (BU) measurement $\mathbf{y}_{\mathrm{BU}}$ that corresponds to the tracked particle. Note that both localization approaches do not

take into account the predicted measurement $\hat{\mathbf{y}}$ and instead exhaustively examine each position of the image to localize the particle.

3.7.1.3 Top-down Localization via Ellipsoidal Measurements

Instead of exhaustively examining each position, the top-down localization scheme in our approach generates measurements in a neighborhood around the predicted measurement $\hat{\mathbf{y}} \in \mathcal{Y}$. In biological imaging (e.g., [64], [178], [98], [116]), typically the measurement space $\mathcal{Y}$ includes variables describing the position $\mathbf{p}$ of the particle in the image as well as variables describing the appearance (e.g., the peak intensity I_{max} of the particle). Certainly, one could generate measurements in the neighborhood of $\hat{\mathbf{y}}$ over the whole space $\mathcal{Y}$. For efficiency reasons, in our approach we only generate position measurements $\mathbf{p}$ close to the predicted position $\hat{\mathbf{p}}$ within the position space $\mathcal{P}$ defined by the image. The predicted position $\hat{\mathbf{p}}$ can be obtained by projecting the predicted measurement $\hat{\mathbf{y}}$ to $\mathcal{P}$ via the projection matrix $\mathbf{\Phi}$:

$$\hat{\mathbf{p}} = \mathbf{\Phi}\hat{\mathbf{y}}. \tag{3.25}$$

To define a neighborhood around $\hat{\mathbf{p}}$, first we define a submatrix $\mathbf{S}_p$ by transforming $\mathbf{S}$ onto $\mathcal{P}$:

$$\mathbf{S}_p = \mathbf{\Phi}\mathbf{S}\mathbf{\Phi}^T. \tag{3.26}$$

Our approach generates measurements based on the predicted position distribution $\mathcal{N}(\cdot; \hat{\mathbf{p}}, \mathbf{S}_p)$. The neighborhood around $\hat{\mathbf{p}}$ is described by the following ellipsoidal validation region $V_{\hat{\mathbf{p}},\mathbf{S}_p}(\gamma_p)$:

$$V_{\hat{\mathbf{p}},\mathbf{S}_p}(\gamma_p) \equiv \{\mathbf{p} \,|\, (\mathbf{p} - \hat{\mathbf{p}})^T \mathbf{S}_p^{-1} (\mathbf{p} - \hat{\mathbf{p}}) \leq \gamma_p^2\}. \tag{3.27}$$

Within this region $V_{\hat{\mathbf{p}},\mathbf{S}_p}(\gamma_p)$ one finds position measurements $\mathbf{p}$ that entail a Mahalanobis distance to $\hat{\mathbf{p}}$ less than or equal to γ_p^2.

Top-down position measurements within the ellipsoidal region $V_{\hat{\mathbf{p}},\mathbf{S}_p}(\gamma_p)$ can be generated via, for instance, proposing split measurements (e.g., [65]), or performing random sampling (e.g., [149], [179]). In our approach we generate measurements by first diagonalizing $\mathbf{S}_p$ thus obtaining the semi-axes of the ellipsoidal region $V_{\hat{\mathbf{p}},\mathbf{S}_p}(\gamma_p)$ as:

$$\mathbf{r}_i = \gamma_p \sqrt{\lambda_i} \mathbf{e}_i, \tag{3.28}$$

where λ_i and $\mathbf{e}_i$ are the eigenvalues and eigenvectors of $\mathbf{S}_p$, respectively. The ellipsoid $E_{\hat{\mathbf{p}},\mathbf{S}_p}$ is explicitly described by its centroid $\hat{\mathbf{p}}$ and its semi-axes $\mathbf{r}_i$. The measurements are generated by taking positions $\mathbf{p}_{j,c}$ along N_c concentric ellipsoidal contours centered

at $\hat{\mathbf{p}}$:

$$\mathbf{p}_{j,c} = \hat{\mathbf{p}} + \frac{c}{N_c}\mathbf{A}\mathbf{u}_j, \tag{3.29}$$

where $\mathbf{u}_j$ is a position along an ellipsoid $E^{\text{canon}}_{\mathbf{0},\mathbf{S}_p}$ in canonical position whose semi-axes have the same magnitude as the semi-axes of $E_{\hat{\mathbf{p}},\mathbf{S}_p}$, $\mathbf{A}$ is a rotation matrix describing the orientation of $E_{\hat{\mathbf{p}},\mathbf{S}_p}$, and $c = 1, 2, \ldots, N_c$ is the concentric index. In 2D images, along a single ellipsoidal contour, we take N_j ellipsoidal positions $\mathbf{u}_j$ that take the following parametric form:

$$\mathbf{u}_j = \begin{pmatrix} |\mathbf{r}_0| \cos \frac{2\pi j}{N_j} \\ |\mathbf{r}_1| \sin \frac{2\pi j}{N_j} \end{pmatrix}, \tag{3.30}$$

where $|\mathbf{r}_0| \geq |\mathbf{r}_1|$, and $j = 1, 2, \ldots, N_j$. The rotation matrix $\mathbf{A}$ is a matrix whose columns are given by the eigenvectors $\mathbf{e}_i$:

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{e}_0 & \mathbf{e}_1 \end{pmatrix}, \tag{3.31}$$

where for the corresponding eigenvalues we have $\lambda_0 \geq \lambda_1$. In 3D images, the measurements $\mathbf{p}_{j,k,c}$ are additionally indexed by the parameter k, and are defined as follows:

$$\mathbf{p}_{j,k,c} = \hat{\mathbf{p}} + \frac{c}{N_c}\mathbf{A}\mathbf{u}_{j,k}. \tag{3.32}$$

Here we take $N_j N_k$ ellipsoidal positions $\mathbf{u}_{j,k}$ which are defined as follows:

$$\mathbf{u}_{j,k} = \begin{pmatrix} |\mathbf{r}_0| \cos \frac{2\pi j}{N_j} \sin \frac{\pi k}{N_k} \\ |\mathbf{r}_1| \sin \frac{2\pi j}{N_j} \sin \frac{\pi k}{N_k} \\ |\mathbf{r}_2| \cos \frac{\pi k}{N_k} \end{pmatrix}, \tag{3.33}$$

where $|\mathbf{r}_0| \geq |\mathbf{r}_1| \geq |\mathbf{r}_2|$, $j = 1, 2, \ldots, N_j$, and $k = 1, 2, \ldots, N_k$. The rotation matrix $\mathbf{A}$ is given as follows:

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{e}_0 & \mathbf{e}_1 & \mathbf{e}_2 \end{pmatrix}, \tag{3.34}$$

where $\lambda_0 \geq \lambda_1 \geq \lambda_2$. Since the measurements are generated along ellipsoidal contours we refer to these measurements as *ellipsoidal measurements*. Our scheme thus generates $N_c N_j N_k$ (in 2D, $N_k = 1$) measurements based on $\mathcal{N}(\cdot; \hat{\mathbf{p}}, \mathbf{S}_p)$ within the position space $\mathcal{P}$. As an example, Figures 3.4 and 3.5 show the ellipsoidal measurements obtained for sample 2D and 3D Gaussian distributions, respectively. The Kalman filter, however, expects measurements $\mathbf{y}$ within $\mathcal{Y}$. For this reason we map the position measurements back to $\mathcal{Y}$. Each i-th measurement $\mathbf{p}_i$ is embedded into $\mathcal{Y}$ via the pseudoinverse $\mathbf{\Phi}^+$:

$$\mathbf{y}_{i,p} = \mathbf{\Phi}^+ \mathbf{p}_i. \tag{3.35}$$

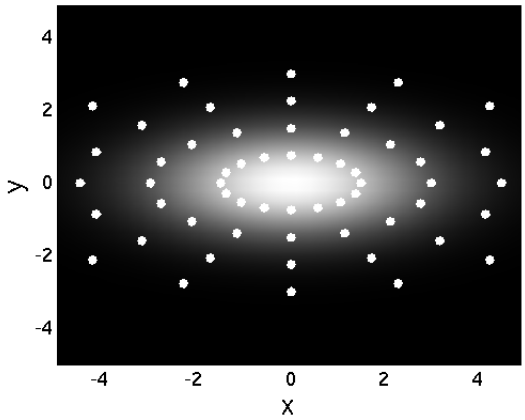


Figure 3.4: Ellipsoidal measurements (dots) for a 2D anisotropic Gaussian distribution. For the Gaussian distribution, brighter intensities correspond to higher probabilities. Here $\gamma_p = 3$. Also $N_c = 4$ and $N_j = 16$; thus 64 measurements are generated.

The vector $\mathbf{y}_{i,p}$ only includes the position information and must be supplemented with the additional variables (e.g., appearance variables) in $\mathcal{Y}$. We extract these variables from the predicted measurement $\hat{\mathbf{y}}$ via a selection matrix $\mathbf{\Psi}$ and add them to the embedded vector $\mathbf{y}_{i,p}$ to obtain measurement $\mathbf{y}_i$:

$$\mathbf{y}_i = \mathbf{y}_{i,p} + \mathbf{\Psi}\hat{\mathbf{y}}. \tag{3.36}$$

Together with the predicted measurement $\hat{\mathbf{y}}$ the top-down localization strategy yields $N_c N_j N_k + 1$ measurements in total.

3.7.1.4 Ellipsoidal Measurements for Bottom-up Measurement

The prediction generated for time step t by the Kalman filter does not take into account the image data of time step t (current image data). Thus the ellipsoidal measurements based on the predicted position distribution $\mathcal{N}(\cdot; \hat{\mathbf{p}}, \mathbf{S}_p)$ do not take into account the current image data. The current bottom-up measurement $\mathbf{y}_{\mathrm{BU}}$ encodes information about the current image data. To take into account the current image data, we generate ellipsoidal measurements based on the measurement distribution $\mathcal{N}(\cdot; \mathbf{p}_{\mathrm{BU}}, \mathbf{R}_p)$, where $\mathbf{p}_{\mathrm{BU}} = \mathbf{\Phi}\mathbf{y}_{\mathrm{BU}}$ includes the position variables of the bottom-up measurement $\mathbf{y}_{\mathrm{BU}}$ and $\mathbf{R}_p = \mathbf{\Phi}\mathbf{R}\mathbf{\Phi}^T$ is a submatrix derived from the covariance matrix $\mathbf{R}$ that regulates the measurement noise process. For embedding the position measurements into the measurement space, we use the bottom-up measurement $\mathbf{y}_{\mathrm{BU}}$ instead of the predicted measurement $\hat{\mathbf{y}}$ in (3.36). Thus, for $\mathcal{N}(\cdot; \mathbf{p}_{\mathrm{BU}}, \mathbf{R}_p)$, together with the bottom-up measurement $\mathbf{y}_{\mathrm{BU}}$, we generate $N_c N_j N_k + 1$ additional measurements. Taking into account the previous $N_c N_j N_k + 1$ measurements generated for $\mathcal{N}(\cdot; \hat{\mathbf{p}}, \mathbf{S}_p)$, our measurement pro-

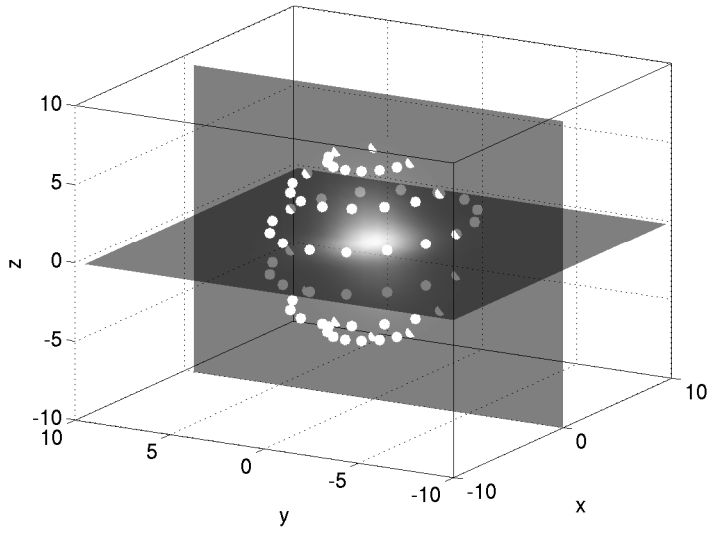


Figure 3.5: Ellipsoidal measurements (dots) for a 3D anisotropic Gaussian distribution. For the Gaussian distribution, brighter intensities correspond to higher probabilities. Here $\gamma_p = 3$. Also $N_c = 1$, $N_j = 16$, and $N_k = 8$; thus 128 measurements are generated.

cess yields up to $N_m = 2N_cN_jN_k + 2$ measurements in total. Note that if the bottom-up measurement $\mathbf{y}_{\mathrm{BU}}$ is missing (e.g., due to image noise), $\mathcal{N}(\cdot; \mathbf{p}_{\mathrm{BU}}, \mathbf{R}_p)$ is undefined and the corresponding ellipsoidal measurements cannot be determined. In that case only half the number of measurements is generated. Thus, regardless of the performance of the bottom-up particle localization scheme, our measurement process supplies the Kalman filter with at least $N_m = N_cN_jN_k + 1$ measurements at each time step.

3.7.1.5 Measurement Integration via Probabilistic Data Association

The probabilistic data association algorithm (e.g., [42], [102]; see Section 2.1.2.2.1) is an approach for solving the problem of motion correspondence, which entails determining one-to-one correspondences between the tracked object and the set of measurements $Y_t \triangleq \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_m}\}$ found within the validation region $V_{\hat{\mathbf{y}},\mathbf{S}}(\gamma)$ (cf. (3.27)). A standard approach for solving this task is to select the measurement $\mathbf{y}_i$ that is closest to the predicted measurement $\hat{\mathbf{y}}$. This local *nearest neighbor* strategy may lead to incorrect correspondences in difficult tracking scenarios (e.g., low SNR), where the set Y_t may miss the true measurement corresponding to the tracked object, and/or may include false measurements. Instead of committing to a single measurement (i.e., hard association), the probabilistic data association algorithm computes an *association probability* β_i to quantify the degree of correspondence between the tracked object and the measurement $\mathbf{y}_i$. To take into account all measurements within the validation region (i.e.,

soft association), a combined innovation $\boldsymbol{\nu}_{\text{comb}}$ is computed:

$$\boldsymbol{\nu}_{\text{comb}} = \sum_{i=1}^{N_m} \beta_i \boldsymbol{\nu}_i \tag{3.37}$$

with $\boldsymbol{\nu}_i = \mathbf{y}_i - \hat{\mathbf{y}}$ and $\sum_{i=0}^{N_m} \beta_i = 1$, where β_0 is the probability that none of the measurements corresponds to the tracked object. The combined innovation $\boldsymbol{\nu}_{\text{comb}}$ is used to perform the update of the prediction in the Kalman filter (see (3.15)). To compute the association probabilities β_i, the standard PDA algorithm assumes that at most one measurement within Y_t represents the true measurement corresponding to the tracked object and that all other measurements are false measurements. It is also assumed that the false measurements are independent and identically distributed (i.i.d.) and that they arise from a uniform distribution over the validation region $V_{\hat{\mathbf{y}},\mathbf{S}}(\gamma)$. Additionally, the number of false measurements is assumed to follow a Poisson distribution. The exact formula for the association probabilities β_i as used in the standard PDA algorithm is given in [102] as:

$$\beta_i = \begin{cases} a_i/C, & i = 1, ..., N_m \\ b/C, & i = 0 \end{cases} \tag{3.38}$$

where

$$a_i \triangleq \exp\left(-\frac{1}{2}\boldsymbol{\nu}_i^T \mathbf{S}^{-1} \boldsymbol{\nu}_i\right) \tag{3.39}$$

$$b \triangleq \lambda_F \left|2\pi\mathbf{S}\right|^{\frac{1}{2}} \frac{1 - P_D P_G}{P_D} \tag{3.40}$$

$$C \triangleq b + \sum_{i=1}^{N_m} a_i. \tag{3.41}$$

Here N_m is the number of measurements, $\mathbf{S}$ is the predicted measurement covariance, $\boldsymbol{\nu}_i$ is the innovation due to the i-th measurement, λ_F represents the expected number of false measurements, P_D denotes the detection probability of true targets, and P_G is the gate probability that arises from restricting the validation region $V_{\hat{\mathbf{y}},\mathbf{S}}(\gamma)$ with the threshold γ^2.

In our approach, because of our ellipsoidal sampling scheme for localization (cf. Section 3.7.1.3), some of the fundamental assumptions of the PDA algorithm are not satisfied. For example, taking Y_t as the set of measurements generated by our measurement process, the 'false' measurements in Y_t do not follow a uniform distribution (but rather all measurements are distributed according to a Gaussian density). In addition, since we generate measurements based on either the predicted measurement $\hat{\mathbf{y}}$ or based on the bottom-up measurement $\mathbf{y}_{\text{BU}}$, the two respective sets of ellipsoidal measurements do

not follow the same distribution. Also, the number of measurements generated by our scheme is constant, so the number of 'false' measurements does not follow a Poisson distribution. A final consideration is that in the standard PDA algorithm the association probabilities only take into account the Mahalanobis distance between each measurement $\mathbf{y}_i$ and the predicted measurement $\hat{\mathbf{y}}$, i.e., the image data is not directly used in the computation of β_i. For these reasons, we adopt a different strategy for computing β_i. Instead of treating β_i as association probabilities, we interpret them as weights that quantify the probability that the image intensities $\mathbf{z}$ conform to the Kalman-based prediction, as represented by the measurements raised with our measurement process. To compute the weights, we query the *image likelihood* $p(\mathbf{z}|\mathbf{x})$, and thus $\beta_i \triangleq p(\mathbf{z}|\mathbf{x} = \mathbf{x}_i)$, where $\mathbf{x}_i = \Xi(\mathbf{y}_i; \hat{\mathbf{x}})$ and the function $\Xi(\cdot;\cdot)$ transforms its argument (measurement $\mathbf{y}$) onto the state space $\mathcal{X}$ taking into account additional variables in the predicted state vector $\hat{\mathbf{x}} = \hat{\mathbf{m}}$. Given the image intensities $\mathbf{z}$ within a region-of-interest (ROI) about the position encoded in $\mathbf{x}$, the image likelihood $p(\mathbf{z}|\mathbf{x})$ is defined by the ratio:

$$p(\mathbf{z}|\mathbf{x}) \triangleq \frac{p_{\mathrm{o}}(\mathbf{z}|\mathbf{x})}{p_{\mathrm{b}}(\mathbf{z}|\mathbf{b})}. \tag{3.42}$$

Here the image likelihood $p_{\mathrm{o}}(\mathbf{z}|\mathbf{x})$ relative to the object is defined as:

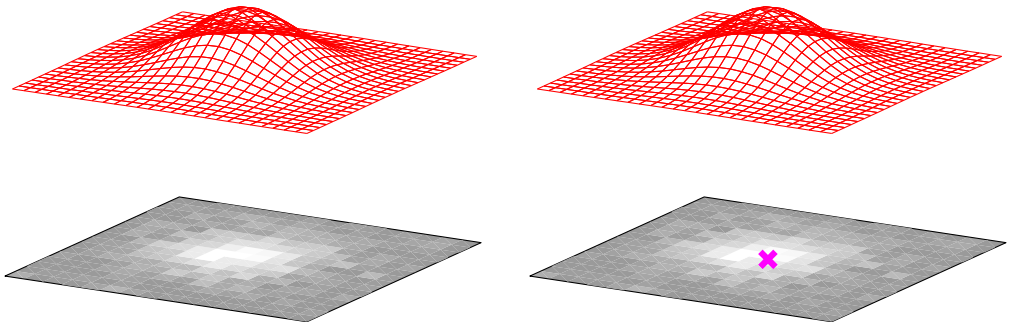
$$p_{\mathrm{o}}(\mathbf{z}|\mathbf{x}) \propto \exp\left(-\frac{D(\mathbf{z}, \mathbf{g}(\mathbf{x}))^2}{2\sigma_n^2}\right), \tag{3.43}$$

where $\mathbf{g}(\cdot)$ represents the image intensities synthesized via a parametric appearance model $g(\cdot)$, $D(\cdot)$ is the Euclidean distance, and σ_n describes the expected degree of noise. The image likelihood $p_{\mathrm{b}}(\mathbf{z}|\mathbf{b})$ relative to the background intensity is given as:

$$p_{\mathrm{b}}(\mathbf{z}|\mathbf{b}) \propto \exp\left(-\frac{D(\mathbf{z}, \mathbf{b})^2}{2\sigma_n^2}\right), \tag{3.44}$$

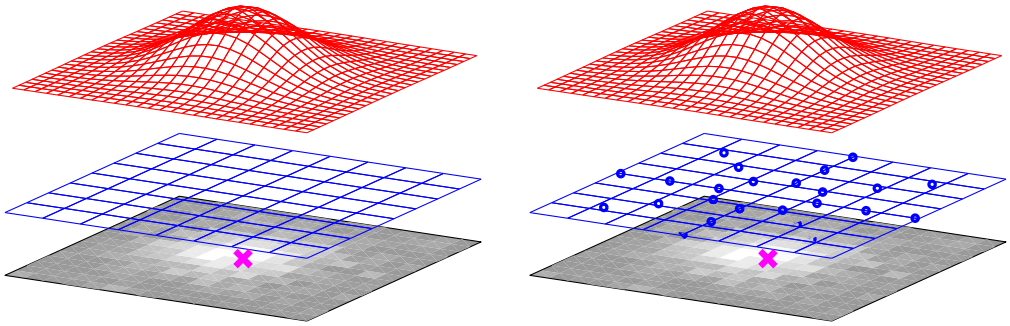
where each element of $\mathbf{b}$ takes the value of the background intensity I_b. Note that a similar image likelihood is used to compute the weights of samples in tracking approaches based on particle filters (e.g., [176], [70]), while here we use a Kalman filter. Once all weights have been evaluated with the image likelihood $p(\mathbf{z}|\mathbf{x})$, the weights $\beta_i, i = 1, \ldots, N_m$ are normalized so that they sum up to unity. Since in our approach the variables β_i are computed using the principle of recognition-by-synthesis, the combined innovation is not used for the problem of motion correspondence (as in the standard PDA algorithm) but rather assimilates multiple measurements into the Kalman filter taking into account the image information directly, i.e., segmentation of a particle from the image data is not necessarily required for estimating the position over time. A

diagram illustrating the measurement process in 2D images is shown in Figure 3.6. In the following, we refer to our approach based on probabilistic data association in conjunction with the ellipsoidal sampling scheme as PDAE.



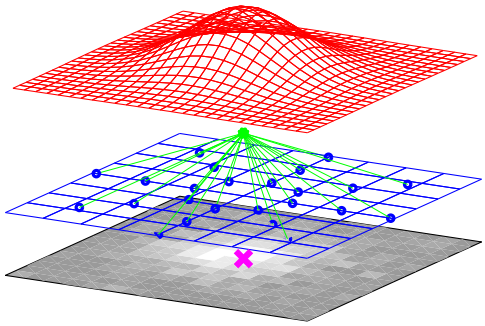
(a) Fluorescent particle and underlying Gaussian representation

(b) Kalman prediction



(c) Exhaustive bottom-up localization

(d) Ellipsoidal measurements



(e) Measurement integration via PDA

Figure 3.6: Proposed measurement process.

3.7.2 Tracking Multiple Fluorescent Particles

To track multiple objects, we instantiate one Kalman filter per object. The measurement process as outlined in Section 3.7.1 considers only one object and therefore must be extended to accommodate the task of tracking multiple fluorescent particles. Concretely, two parts of the measurement process need additional steps: first, for the bottom-up localization step in Section 3.7.1.2, we assume that only one bottom-up measurement is generated by the spot detection scheme. This assumption is not satisfied within the context of tracking multiple fluorescent particles, and thus a motion correspondence step is required. We use the motion correspondence step presented in Section 3.3 above. Second, we assume that modes in the image likelihood $p(\mathbf{z}|\mathbf{x})$ used to evaluate the probabilities β in the measurement integration step in Section 3.7.1.5 are caused by the tracked object only. Within the context of tracking multiple fluorescent particles, such modes may actually be caused by neighboring objects with a similar appearance. To minimize the influence of modes induced by other objects, we calculate the support of an image position relative to the neighboring objects of each tracked object (see Section 3.7.2.2 below).

Note that since we use the combined innovation principle of PDA in our measurement process, it would be conceivable that the task of tracking multiple objects could be addressed via the joint probabilistic data association (JPDA) algorithm (see Section 2.1.2.2.1). In the JPDA algorithm, each association probability β is calculated over all possible *global* association *hypotheses* between the tracked objects and the measurements. This requires deterministically enumerating all possible global association hypotheses. Since our measurement scheme generates a large number of measurements, the JPDA algorithm would entail a high computational cost. Additionally, in the JPDA algorithm, the same assumptions concerning false measurements are made as in the PDA algorithm. As mentioned above, in our measurement process such fundamental assumptions are not met (cf. Section 3.7.1.5). The JPDA algorithm also assumes a one-to-one correspondence between the tracked objects and the measurements. However, in our measurement process, the measurements generated with our ellipsoidal sampling scheme are already associated to the corresponding object (filter) and thus we have a many-to-one correspondence. Finally, we do not interpret the weight β as an association probability but rather as a weight computed based on an image likelihood. For these reasons, the JPDA algorithm is not well suited to deal with the task of tracking multiple particles within the context of our measurement process.

3.7.2.1 Motion Correspondence

We use the motion correspondence step presented in Section 3.3 above. After finding one-to-one correspondences, predicted measurements (i.e., and thus filters) in $\hat{Y}$ may remain unmatched. In this case, the measurement process of that filter generates only measurements based on the predicted measurement.

3.7.2.2 Image Support Relative to Neighbors

In our measurement process, the probabilities β are calculated by querying the image likelihood $p(\mathbf{z}|\mathbf{x})$. This likelihood is constructed under the assumption that only one object is visible in the image, i.e., independently from other objects. Under this assumption, a mode (or peak) in the image likelihood corresponds to the tracked object. Within the context of tracking multiple objects, a peak in the image likelihood may correspond to a neighboring object with a similar appearance. Thus, there is a risk that a peak originating from a neighboring object may strongly influence the estimates of the Kalman filter, eventually leading the filter to converge on neighboring objects. To prevent this, the calculation of the probabilities β needs to also take into consideration the probabilities β of the neighboring filters. After calculating the corresponding image likelihoods of all filters, each filter j is associated with a set of weighted position measurements $\{\mathbf{p}_k^j = \mathbf{\Phi}\mathbf{y}_k^j, \beta_k^j\}_{k=1}^{N_m}$. Likewise, the neighboring filters $Nb(j)$ around filter j allow constructing a set of measurements $\cup_{i\in Nb(j)}\{\mathbf{p}_k^i, \beta_k^i\}$. By considering all positions $\mathbf{p}$ up to pixel accuracy we define a neighborhood map:

$$M_{Nb(j)}(\mathbf{p}) = \sum_{i\in Nb(j)} \sum_{l\in K_{\mathbf{p}}^i} \beta_l^i, \tag{3.45}$$

where $K_{\mathbf{p}}^i = \{k \,|\, \mathbf{p}_k^i = \mathbf{p}\}$ is the set of indices of measurements in $\cup_{i\in Nb(j)}\{\mathbf{p}_k^i, \beta_k^i\}$ located at $\mathbf{p}$. The map $M_{Nb(j)}$ describes the support of each image position to neighboring filters. The support provided to neighboring objects $Nb(j)$ directly influences the support provided to the j-th tracked object: The higher the support to the neighbors, the lower the support to the j-th tracked object. Thus each position $\mathbf{p}$ allows for a relative support $S_j(\mathbf{p}) = \exp(-M_{Nb(j)}(\mathbf{p}))$ for object j. For reasons of robustness, we also take into account neighboring positions to compute the support $S_j(\mathbf{p})$ at position $\mathbf{p}$:

$$S_j(\mathbf{p}) = \exp\left(\sum_{\mathbf{p}'} M_{Nb(j)}(\mathbf{p}')\log(\theta(\mathbf{p},\mathbf{p}'))\right) \tag{3.46}$$

where $\theta(\mathbf{p}, \mathbf{p}')$ describes the spatial dependency between two image positions in general and is defined as follows [220]:

$$\theta(\mathbf{p}, \mathbf{p}') = 1 - \exp\left(-\frac{D(\mathbf{p}, \mathbf{p}')^2}{\sigma_p^2}\right). \tag{3.47}$$

In (3.46), $\log(\theta(\mathbf{p}, \mathbf{p}'))$ acts as a weighting factor over the neighboring positions. The parameter σ_p regulates the strength of the contribution of neighboring positions towards the calculation of the support. Since for the same position $\mathbf{p}$ we have $\theta(\mathbf{p}, \mathbf{p}) = 0$ and thus $\log(\theta(\mathbf{p}, \mathbf{p})) = -\infty$, we use a constant weight in this case (in our algorithm we use a weight of -5). The probability $\hat{\beta}_k^j$ for a measurement $\mathbf{y}_k^j$ of object j is computed as:

$$\hat{\beta}_k^j = p(\mathbf{z}|\mathbf{x} = \Xi(\mathbf{y}_k^j; \hat{\mathbf{x}}))S_j(\mathbf{\Phi}\mathbf{y}_k^j), \tag{3.48}$$

and $\hat{\mathbf{x}} = \hat{\mathbf{m}}$ is the prediction of the Kalman filter (3.9). The weights $\hat{\beta}_k^j$ are subsequently normalized so that they sum up to unity. This process is performed for all objects. Since the probabilities $\hat{\beta}$ of the neighbors are eventually different from the original probabilities β that were used to calculate the support, we recalculate the neighborhood map using the new weights $\hat{\beta}$ and recalculate the probabilities. We perform this process (i.e., recalculating the neighborhood map and calculating the probabilities) a certain number of times. In practice we found that the values of the weights become relatively stable after five iterations over all objects.

3.7.3 Multiple Motion Models

To better describe the motion of particles that alternate between distinct motion patterns, we use the interacting multiple model (IMM) algorithm (e.g., [21], [22], [64]). Below we present the IMM algorithm. We also introduce an approach for combining our PDAE approach with the IMM algorithm.

3.7.3.1 Interacting Multiple Model Algorithm

The standard IMM algorithm assumes that the motion of a particle is described by a finite set Ω of N_Ω motion models. One Kalman filter is instantiated per motion model $\omega \in \Omega$. Additionally, it is assumed that the dynamical model of a particle is well described by a Markov chain with transition matrix $\mathbf{\Delta}$. The IMM algorithm recursively computes a model-conditioned posterior density $p(\mathbf{x}_t|\omega_t, \mathbf{y}_{1:t})$ as well as a posterior model probability $P(\omega_t|\mathbf{y}_{1:t})$. One iteration of the IMM algorithm involves the following steps: First, the algorithm computes a predicted probability $\hat{P}_i \equiv P(\omega_t = i|\mathbf{y}_{1:t-1})$ of each

model $\omega = i$ as follows:

$$\hat{P}_i \equiv P(\omega_t = i|\mathbf{y}_{1:t-1}) = \sum_{j=1}^{N_\Omega} P(\omega_{t-1} = j|\mathbf{y}_{1:t-1})\delta_{ij}, \tag{3.49}$$

where δ_{ij} is an element of the transition matrix $\mathbf{\Delta}$. Using the predicted probability $\hat{P}_i$ of each model $\omega = i$, the algorithm calculates the *mixing* probabilities $P_{j\to i}$ as follows:

$$P_{j\to i} \equiv \frac{\delta_{ij} P(\omega_{t-1} = j|\mathbf{y}_{1:t-1})}{\hat{P}_i}. \tag{3.50}$$

For each i-th model, the IMM algorithm computes a *mixed* state and covariance matrix taking into account the mixing probabilities $P_{j\to i}$ as well as the mean vector $\mathbf{m}_{t-1}^j$ and covariance matrix $\mathbf{P}_t^j$ of each j-th Kalman filter that were previously estimated:

$$\mathbf{m}_{t-1}^{i,\text{mix}} = \sum_{j=1}^{N_\Omega} P_{j\to i}\mathbf{m}_{t-1}^j \tag{3.51}$$

$$\mathbf{P}_{t-1}^{i,\text{mix}} = \sum_{j=1}^{N_\Omega} P_{j\to i}\left\{\mathbf{P}_{t-1}^j + \left[\mathbf{m}_{t-1}^j - \mathbf{m}_{t-1}^{i,\text{mix}}\right]\left[\mathbf{m}_{t-1}^j - \mathbf{m}_{t-1}^{i,\text{mix}}\right]^T\right\}. \tag{3.52}$$

The mixing probabilities $P_{j\to i}$ quantify the relative contribution of each j-th model towards the initial condition of the i-th model. The mixed state $\mathbf{m}_{t-1}^{i,\text{mix}}$ and covariance $\mathbf{P}_{t-1}^{i,\text{mix}}$ are used as input for the i-th Kalman filter (see Section 3.5). Based on the corresponding mixed estimate and covariance, each i-th model-matched filter performs a prediction using the corresponding dynamical model $\mathbf{F}^i$ as well as the covariance matrix $\mathbf{Q}^i$ (see (3.9) as well as (3.10)). Taking into account a measurement $\mathbf{y}_t$ obtained from the images, each filter is updated using the standard Kalman filter equations, thereby obtaining the model-conditioned posterior $p(\mathbf{x}_t|\omega_t, \mathbf{y}_{1:t})$. The IMM algorithm also updates the predicted probabilities of each dynamical model ω via Bayes' theorem:

$$P(\omega_t = i|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})\,\hat{P}_i}{\sum_j p(\mathbf{y}_t|\omega_t = j, \mathbf{y}_{1:t-1})\,\hat{P}_j} \tag{3.53}$$

where the likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model is given by evaluating the predicted measurement density $\mathcal{N}(\cdot; \hat{\mathbf{y}}^i, \mathbf{S}^i)$ (which is calculated by the i-th Kalman filter) with the measurement $\mathbf{y}_t$ obtained from the images. In other words, the likelihood of each model is given by a Gaussian density. As output, the algorithm calculates a weighted estimate over all N_Ω Kalman filters, where the weights are given by the

posterior model probabilities (see (3.53)):

$$\mathbf{m}_t = \sum_{j=1}^{N_\Omega} P(\omega_t = j|\mathbf{y}_{1:t})\mathbf{m}_t^j \tag{3.54}$$

$$\mathbf{P}_t = \sum_{j=1}^{N_\Omega} P(\omega_t = j|\mathbf{y}_{1:t}) \left\{ \mathbf{P}_t^j + \left[\mathbf{m}_t^j - \mathbf{m}_t\right] \left[\mathbf{m}_t^j - \mathbf{m}_t\right]^T \right\}. \tag{3.55}$$

3.7.3.2 Combining the PDAE approach with the IMM Algorithm

Our measurement process as outlined in Section 3.7.1 above generates multiple measurements for a single Kalman filter. The measurements are integrated into a single Kalman filter via the combined innovation principle of the PDA algorithm (see Section 3.7.1.5). Instead of considering a single Kalman filter per object, the IMM algorithm instantiates N_Ω model-matched Kalman filters per object. However, in the standard IMM algorithm, only a *single* measurement $\mathbf{y}_t$ is used to update all N_Ω filters. Instead of relying on a single measurement, we use the PDAE approach to feed *multiple* measurements to each filter: each i-th Kalman filter raises its own set $\{\mathbf{y}_k, \beta_k, \omega_k = i\}_{k=1}^{N_m}$ of ellipsoidal measurements. Note that besides being associated with an association probability β, each measurement is also associated with the motion model $\omega_k = i$ that led to such a measurement. In the standard IMM algorithm, the likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model is given by a Gaussian density. In the standard PDA-IMM extension (e.g., [102]), the probability of each measurement associated with the i-th filter is evaluated by querying the predicted measurement density $\mathcal{N}(\cdot; \hat{\mathbf{y}}^i, \mathbf{S}^i)$. The mean value of these Gaussian likelihoods over all measurements is used to calculate the likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model. In our case, to calculate the likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model, we first take the unnormalized values of the weights β_k of each i-th filter and normalize them with respect to the sum of all weights of all N_Ω filters:

$$\beta'_k = \frac{\beta_k}{\sum_c^{N_\Omega} \sum_{j \in \Upsilon_c} \beta_j}, k \in \Upsilon_i \tag{3.56}$$

where the set $\Upsilon_i = \{j \,|\, \beta_j, \omega_j = i\}$ indexes the weights of the measurements of the i-th filter. The likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model is given as the sum over its normalized weights β'_k:

$$p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1}) = \sum_{j \in \Upsilon_i} \beta'_j \tag{3.57}$$

Since the weights β reflect the degree to which the image intensities conform to the predictions of each filter, the likelihood $p(\mathbf{y}_t|\omega_t = i, \mathbf{y}_{1:t-1})$ of each i-th model may be seen as the probability that the image intensities conform to the predictions generated

by the i-th model *relative* to the probability that the image intensities conform to the predictions generated by the other models. This is in contrast to the standard PDA-IMM extension, where the likelihood of each model does not take into account the likelihoods of the other models.

When tracking multiple objects, challenges arise in the motion correspondence step (see Section 3.3) as well as in the calculation of the image support (see Section 3.7.2.2). In the motion correspondence step, we assume that each tracked object is represented by at most a *single* predicted measurement $\hat{\mathbf{y}}$. The IMM algorithm instead uses N_Ω predicted measurements to represent an object. While the predicted measurements could be mixed using the mixing probabilities $P_{j\to i}$ (e.g., [113]) to obtain a mean prediction, this could lead to a less accurate prediction, since the mean prediction does not exactly represent any particular motion model. Another scheme [64] uses all N_Ω predicted measurements of all N_{obj} objects as input for the correspondence scheme. There a (local) nearest neighbor (see Section 2.1.1.2.1) is used to sequentially assign $N_{m,\text{BU}}$ bottom-up measurements to all $N_\Omega N_{\text{obj}}$ predicted measurements. Once a measurement is assigned to a predicted measurement of a filter corresponding to a certain object, the other $N_\Omega - 1$ filters corresponding to the object are removed from the correspondence problem. This greedy strategy may lead to errors in cases where objects are in close proximity. In our approach, for each object we select the filter with the highest predicted probability $\hat{P}_i$ (3.49). This strategy leads to N_{obj} predicted measurements and so the correspondence scheme based on a global nearest neighbor strategy (see Section 3.3) remains applicable. The measurement assigned to the filter with the highest predicted probability of each object is afterwards shared with all other $N_\Omega - 1$ filters of the object. For the calculation of the image support (see Section 3.7.2.2), we treat all filters $N_\Omega N_{\text{obj}}$ as independent filters. This allows the N_Ω filters of a certain object to compete for image positions among themselves, which helps to select the best motion model among all N_Ω motion models relative to the predicted positions of neighboring objects.

3.8 Models for Tracking Fluorescent Particles

In this section, we describe the appearance, dynamical, and measurement models that we have used in our tracking approaches. We describe models for round particles as well as for elongated particles.

3.8.1 Round Particles with a Random Walk Motion Model

In the following, we define models for tracking particles with a round appearance exhibiting random motion. The models are particularly suitable for tracking virus particles.

Models for 2D as well as 3D images are presented.

3.8.1.1 Appearance Model for Round Particles

In our approach, the intensities of each particle are represented by a 2D or 3D Gaussian function. The Gaussian function is parametrized by the position $\mathbf{p}_0 = (x_0, y_0)^T$ or $\mathbf{p}_0 = (x_0, y_0, z_0)^T$ of the particle in a 2D or 3D image, respectively, by the peak intensity $I_{\max}$, as well as by the standard deviations σ_{xy} (2D case) or σ_{xy}, σ_z (3D case). These parameters are encoded in the state vector $\mathbf{x} = (x_0, y_0, I_{\max}, \sigma_{xy})^T$ (2D case) or $\mathbf{x} = (x_0, y_0, z_0, I_{\max}, \sigma_{xy}, \sigma_z)^T$ (3D case). For 2D images, the appearance model of a single particle is given by:

$$g_{\mathrm{G2D}}(x, y; \mathbf{x}) = I_b + (I_{\max} - I_b) \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma_{xy}^2}\right), \tag{3.58}$$

where I_b denotes the background intensity. For 3D images, the appearance model is described by:

$$g_{\mathrm{G3D}}(x, y, z; \mathbf{x}) = I_b + (I_{\max} - I_b) \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma_{xy}^2} - \frac{(z - z_0)^2}{2\sigma_z^2}\right). \tag{3.59}$$

3.8.1.2 Dynamical Model: Random Motion

In our application, we observed that the motion of the virus particles is well characterized by a random walk model. We assume that the position parameters x_0, y_0 (2D case) or x_0, y_0, z_0 (3D case) as well as the appearance parameters $I_{\max}, \sigma_{xy}$ (2D case) or $I_{\max}, \sigma_{xy}, \sigma_z$ (3D case) follow Gaussian random walks with small perturbations. In 2D, the dynamical model is encoded by the transition matrix $\mathbf{F} = \mathrm{diag}(1, 1, 1, 1)$ and the covariance matrix $\mathbf{Q} = \mathrm{diag}(q_x, q_y, q_{I_{\max}}, q_{\sigma_{xy}})$. In 3D, the transition matrix takes the following form: $\mathbf{F} = \mathrm{diag}(1, 1, 1, 1, 1, 1)$, while the covariance matrix is defined by: $\mathbf{Q} = \mathrm{diag}(q_x, q_y, q_z, q_{I_{\max}}, q_{\sigma_{xy}}, q_{\sigma_z})$. The diagonal elements of $\mathbf{Q}$ account for the expected squared deviations of the corresponding variables over a time interval.

3.8.1.3 Measurement Model for the Kalman Filter and PDAE

We have measurements $\mathbf{y} = (x', y', I'_{\max}, \sigma'_{xy})^T$ (2D case) or $\mathbf{y} = (x', y', z', I'_{\max}, \sigma'_{xy}, \sigma'_z)^T$ (3D case). In 2D, the measurement matrix is $\mathbf{H} = \mathrm{diag}(1, 1, 1, 1)$ while the uncertainties of the measurements are $\mathbf{R} = \mathrm{diag}(r_x, r_y, r_{I_{\max}}, r_{\sigma_{xy}})$. In 3D, the measurement matrix is defined as $\mathbf{H} = \mathrm{diag}(1, 1, 1, 1, 1, 1)$ while the covariance matrix is given as $\mathbf{R} = \mathrm{diag}(r_x, r_y, r_z, r_{I_{\max}}, r_{\sigma_{xy}}, r_{\sigma_z})$. The elements of the diagonal of $\mathbf{R}$ represent the expected squared errors for the corresponding variables. Since the measurement space and

the state space share the same basis, the function $\Xi(\mathbf{y}; \hat{\mathbf{x}})$ that maps the measurements onto the state space is given by:

$$\Xi(\mathbf{y}; \hat{\mathbf{x}}) = \mathbf{I}\mathbf{y}, \tag{3.60}$$

where $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{\Phi}$ used to project a 2D measurement $\mathbf{y}$ onto a 2D position space is defined as follows:

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{3.61}$$

with the pseudoinverse $\mathbf{\Phi}^+ = \mathbf{\Phi}^T$. In 3D, the projection matrix is given by:

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.62}$$

The selection matrix $\mathbf{\Psi}$ in the 2D case is given by $\mathbf{\Psi} = \mathrm{diag}(0, 0, 1, 1)$ while in 3D this matrix takes the form $\mathbf{\Psi} = \mathrm{diag}(0, 0, 0, 1, 1, 1)$.

3.8.1.4 Measurement Model for the Particle Filter

For the approaches based on the particle filter, given the state $\mathbf{x} = (x_0, y_0, I_{max}, \sigma_{xy})^T$, the measurement $\mathbf{y}$ is defined as the image region enclosed via a bounding box centered at (x_0, y_0) with dimensions $(a\lceil\sigma_{xy}\rceil + 1) \times (a\lceil\sigma_{xy}\rceil + 1)$, where $a = 7$ and $\lceil\cdot\rceil$ denotes the ceiling operator; note that the measurement depends on the state. In 3D the state is given by $\mathbf{x} = (x_0, y_0, z_0, I_{\max}, \sigma_{xy}, \sigma_z)^T$, and the measurement $\mathbf{y}$ is given by the image intensities in a region defined by a box centered at (x_0, y_0) with dimensions $(a\lceil\sigma_{xy}\rceil + 1) \times (a\lceil\sigma_{xy}\rceil + 1) \times (a\lceil\sigma_z\rceil + 1)$. The measurement model for both the 2D and 3D cases is defined as the ratio:

$$p(\mathbf{y}|\mathbf{x}) \triangleq \frac{p_{\mathrm{o}}(\mathbf{y}|\mathbf{x})}{p_{\mathrm{b}}(\mathbf{y}|\mathbf{b})}. \tag{3.63}$$

The likelihood $p_{\mathrm{o}}(\mathbf{y}|\mathbf{x})$ relative to the object is the same as in (3.43) and likewise the likelihood $p_{\mathrm{b}}(\mathbf{y}|\mathbf{b})$ relative to the background is defined as in (3.44), where in both cases we have $\mathbf{z} = \mathbf{y}$.

3.8.1.5 Multi-Channel Measurement Model

The non-linear measurement model in (3.63) allows us to deal with two-channel microscopy images comprising a 'red' channel and a 'green' channel. This is achieved analogously to measurement models employed for color images (e.g., RGB images). The proposed measurement model $p_{\text{2-channel}}(\mathbf{y}|\mathbf{x})$ is a straightforward extension of the above-

described one-channel case. The model compares the hypothesized state vector $\mathbf{x}$ with information in both red $\mathbf{y}_r$ and green $\mathbf{y}_g$ channels as follows:

$$p_{\text{2-channel}}(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{D(\mathbf{y}_r, \mathbf{g}(\mathbf{x}))^2 + D(\mathbf{y}_g, \mathbf{g}(\mathbf{x}))^2}{2\sigma_n^2}\right). \tag{3.64}$$

An advantage of this approach for the analysis of two-channel microscopy images is that we do not have to carry out an explicit matching step between the images (as in, e.g., [39]). Extension to m channels is straightforward:

$$p_{\text{m-channel}}(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{\sum_m D(\mathbf{y}_m, \mathbf{g}(\mathbf{x}))^2}{2\sigma_n^2}\right). \tag{3.65}$$

Similarly, this approach can be adapted to accommodate differences in the channels (e.g., one channel may display only the cellular autofluorescence while the other one may display only particles).

3.8.2 Elongated Particles with a Directed Motion Model

Here we present models for tracking particles with an elongated appearance exhibiting directed motion. The models are particularly suitable for tracking microtubule tips. Models for 2D images are presented.

3.8.2.1 Appearance Model for Elongated Particles

For 2D image data of elongated objects, the underlying 2D Gaussian representation takes into consideration the orientation of the object. We represent the object with a 2D anisotropic Gaussian function that is parametrized by the position $\mathbf{p}_0 = (x_0, y_0)^T$, the velocity $\dot{\mathbf{p}} = (\dot{x}, \dot{y})^T$, the peak intensity $I_{\max}$, and the standard deviations σ_x and σ_y. The parameters are represented in the state vector $\mathbf{x} = (x_0, \dot{x}, y_0, \dot{y}, I_{\max}, \sigma_x, \sigma_x)^T$. The appearance of an elongated object is given by:

$$g_{\text{G2DA}}(x, y; \mathbf{x}) = I_b + (I_{\max} - I_b)\, \exp\left(-\frac{1}{2}(\mathbf{p} - \mathbf{p}_0)^T \mathbf{\Sigma}_{\text{G2DA}}^{-1}(\mathbf{p} - \mathbf{p}_0)\right), \tag{3.66}$$

where $\mathbf{p} = (x, y)^T$ and $\mathbf{\Sigma}_{\text{G2DA}} = \mathbf{ADA}^T$ is the covariance matrix defined by the rotation matrix:

$$\mathbf{A} = \begin{pmatrix} \cos(\alpha_{\text{G2DA}}) & -\sin(\alpha_{\text{G2DA}}) \\ \sin(\alpha_{\text{G2DA}}) & \cos(\alpha_{\text{G2DA}}) \end{pmatrix} \tag{3.67}$$

and the diagonal matrix:

$$\mathbf{D} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}. \tag{3.68}$$

The angle α_{G2DA} is given by the velocity components:

$$\alpha_{\text{G2DA}} = \tan\left(\frac{\dot{y}}{\dot{x}}\right). \tag{3.69}$$

This entails that the orientation of the object is coupled with the orientation of the velocity vector of the object. This might limit the capabilities of the model, since the orientation of the object might not be governed by its motion. In the case of microtubule tips, however, the orientation of the tips often agrees with the orientation of the velocity vector. Also coupling the orientation with the motion reduces the number of parameters.

3.8.2.2 Dynamical Model: Directed Motion

In our application, we consider tracking microtubuli tips which belong to the cytoskeleton. The cytoskeleton is a system of filaments providing spatial (e.g., cell organization) and mechanical (e.g., cell growth) functions to the cell. Microtubules are a type of cytoskeletal filaments that guide the position of organelles as well as the intracellular transport process [3]. Studying the dynamic behavior of microtubules allow observing cell formation processes as well as intracellular transport processes. Often the ends (tips) of the microtubules are fluorescently labeled and imaged. Microtubule tips exhibit a strong direction in their movement. We therefore assume a directed motion model with constant velocity for the position $\mathbf{p}_0$ with small deviations for the velocity $\dot{\mathbf{p}}$. The appearance parameters $I_{\max}, \sigma_x, \sigma_y$ are assumed to follow Gaussian random walks with small perturbations. The dynamical model is defined as:

$$\mathbf{F} = \begin{pmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.70}$$

where Δt is the time interval between two time steps. The covariance matrix regulating the perturbations is defined as:

$$\mathbf{Q} = \begin{pmatrix} q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 & 0 & 0 \\ q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{I_{\max}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{\sigma_x} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{\sigma_y} \end{pmatrix}, \tag{3.71}$$

where q_{CV} is the expected squared deviation of the velocity over a time interval Δt.

3.8.2.3 Measurement Model for the Kalman Filter and PDAE

For the Kalman filter and the PDAE we have measurements $\mathbf{y} = (x', y', I'_{\max}, \sigma'_x, \sigma'_y)^T$ and so the measurement matrix is given by:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.72}$$

The uncertainties of the measurements are $\mathbf{R} = \mathrm{diag}(r_x, r_y, r_{I_{\max}}, r_{\sigma_x}, r_{\sigma_y})$. Here the state space and the measurement space do not share the same basis so the function $\Xi(\cdot;\cdot)$ is given by:

$$\Xi(\mathbf{y}; \hat{\mathbf{x}}) = \mathbf{H}^{+}\mathbf{y} + \mathbf{\Lambda}\hat{\mathbf{x}} \tag{3.73}$$

where $\mathbf{H}^{+} = \mathbf{H}^T$ and the matrix $\mathbf{\Lambda} = \mathrm{diag}(0, 1, 0, 1, 0, 0, 0)$ selects the velocity components represented in the prediction $\hat{\mathbf{x}}$. The matrix $\mathbf{\Phi}$ used to project a measurement $\mathbf{y}$ onto a 2D position space is defined as follows:

$$\mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{3.74}$$

The selection matrix $\mathbf{\Psi}$ for the measurements is given by $\mathbf{\Psi} = \mathrm{diag}(0, 0, 1, 1, 1)$. To obtain initial values for the velocity components we use the displacement vectors calculated by the Lucas-Kanade optical flow approach [119] using the image data of two consecutive time steps.

3.8.2.4 Circular Sector Measurements for Dynamic Instability

Microtubule ends (tips) undergo periods of rapid growth as well as periods of disassembly wherein the microtubule ends appear to shrink. This phenomenon is called *dynamic instability* [3]. Dynamic instability results in rapid acceleration and deceleration periods for the microtubule tips. These changes in acceleration may be captured by increasing the values of the velocity parameter q_{CV} but using a too large value for this parameter may lead to tracking errors since changes in acceleration are allowed in all directions. Instead, we sample positions along the direction of the velocity vector $\dot{\mathbf{p}}$ to cope with rapid growth periods. To cope with disassembly, we sample positions around the opposite direction $-\dot{\mathbf{p}}$ of the velocity vector.

Given the predicted position $\hat{\mathbf{p}}$ and the predicted velocity vector $\dot{\mathbf{p}}$ we define a circle centered at $\hat{\mathbf{p}}$ with radius $\mathbf{r}_{\text{cs}} = \gamma_{\text{cs}}|\dot{\mathbf{p}}|$, where γ_{cs} is a factor regulating the size of the circle. We define an arc between positions $\hat{\mathbf{p}} + \mathbf{A}(-\alpha_{\text{cs}}(|\mathbf{r}_{\text{cs}}|))\mathbf{r}_{\text{cs}}$ and $\hat{\mathbf{p}} + \mathbf{A}(\alpha_{\text{cs}}(|\mathbf{r}_{\text{cs}}|))\mathbf{r}_{\text{cs}}$ where $\mathbf{A}(\alpha_{\text{cs}})$ is a rotation matrix and $\alpha_{\text{cs}}(|\mathbf{r}_{\text{cs}}|)$ is a function determining the angle of rotation based on the magnitude of the radius $\mathbf{r}_{\text{cs}}$, which is proportional to the velocity vector. Concretely we have:

$$\mathbf{A}(\alpha_{\text{cs}}) = \begin{pmatrix} \cos(\alpha_{\text{cs}}) & -\sin(\alpha_{\text{cs}}) \\ \sin(\alpha_{\text{cs}}) & \cos(\alpha_{\text{cs}}) \end{pmatrix}, \tag{3.75}$$

and

$$\alpha_{\text{cs}}(|\mathbf{r}_{\text{cs}}|) \begin{cases} \pi/3 & |\mathbf{r}_{\text{cs}}| < 1 \\ \pi/4 & 1 \leq |\mathbf{r}_{\text{cs}}| < 7.5 \\ \pi/16 & 7.5 \leq |\mathbf{r}_{\text{cs}}| \end{cases} \tag{3.76}$$

We obtain $N_{\text{cs},c}(N_{\text{cs},j} + 1)$ measurements over the circular sector (CS) confined by the arc as follows:

$$\mathbf{p}_{\text{CS},j,c} = \hat{\mathbf{p}} + \frac{c}{N_{\text{CS},c}}\mathbf{A}\left(\alpha_{\text{cs}}(|\mathbf{r}_{\text{cs}}|)\left[\frac{2j}{N_{\text{CS},j}} - 1\right]\right)\mathbf{r}_{\text{cs}}, \tag{3.77}$$

where $j = 0, 1, 2, \ldots, N_{\text{cs},j}$ and $c = 1, 2, 3, \ldots, N_{\text{cs},c}$. Figure 3.7 shows a schematic representation of the circular sector measurements. We sample CS measurements in the opposite direction by applying the scheme to $-\mathbf{r}_{\text{cs}}$. We therefore compute $N_{\text{cs},m} = 2(N_{\text{cs},c}(N_{\text{cs},j} + 1))$ circular sector measurements. The rationale for introducing a dependency between the magnitude of the velocity vector $|\mathbf{r}_{\text{cs}}|$ and the angle α_{cs} that regulates the size of the sector is that at high speeds the likelihood of a sudden change in direction is low. Conversely, at low speeds, the likelihood that the object may change its direction is high. Having a larger angle at low speeds allows us to capture such changes in the direction. The CS measurements in the opposite direction serve as an 'anchor' and allow

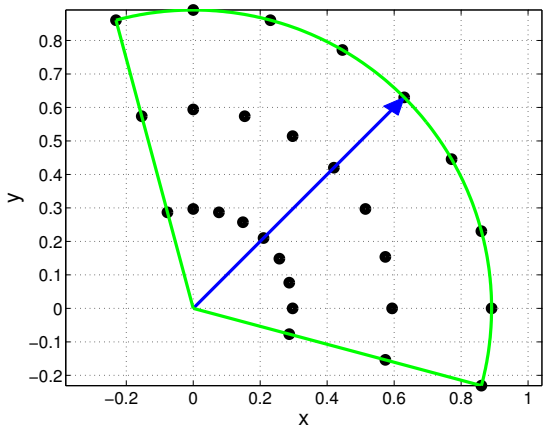
us to cope with sudden decelerations of the object.

Analogous to the ellipsoidal measurements (see Section 3.7.1.3), we evaluate the image likelihood $p(\mathbf{z}|\mathbf{x})$ of the CS measurements by embedding them into the measurement space as well as the state space via $\mathbf{\Psi}^+$ and $\hat{\mathbf{y}}$ (see (3.35)), as well as $\Xi(\cdot;\cdot)$ (see (3.73)), respectively. The measurements are then integrated into the Kalman filter via the combined innovation (see 3.37) that, *in addition*, integrates the ellipsoidal measurements that are raised as by the standard procedure for the PDAE approach. Certainly, there may be an overlap between the two sets of measurements, however, the redundancy emphasizes the importance of certain spatial regions that have a high (a priori) probability for the object. The inclusion of the circular sector measurements shows also the versatility of the PDAE approach for including measurements raised taking into consideration different information or a priori knowledge.

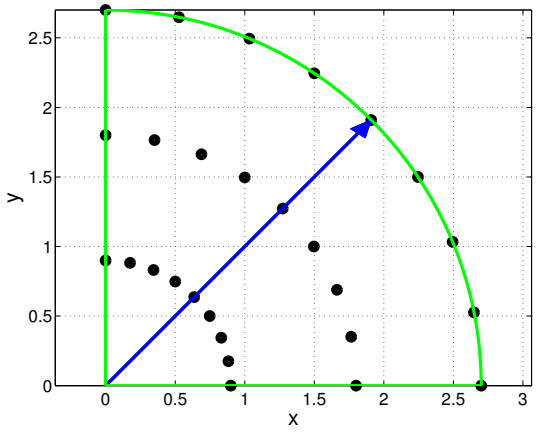
3.8.3 Round Particles with Multiple Motion Models

In cases where round particles alternate between random motion and directed motion, we use the PDAE approach in conjunction with the IMM algorithm (see Section 3.7.3.2) using a random walk model as well as a directed motion model with constant velocity. As appearance model we use the isotropic Gaussian model presented in Section 3.8.1.1. The state vector is given by $\mathbf{x} = (x_0, \dot{x}, y_0, \dot{y}, I_{\max}, \sigma_{xy})^T$ (2D case) or $\mathbf{x} = (x_0, \dot{x}, y_0, \dot{y}, z_0, \dot{z}, I_{\max}, \sigma_{xy}, \sigma_z)^T$ (3D case). For the random walk model ($\omega = 1$), the dynamical model is given in 2D by $\mathbf{F}^1 = \mathrm{diag}(1,1,1,1,1,1)$ and the covariance matrix is $\mathbf{Q}^1 = \mathrm{diag}(q_x, q_{\dot{x}}, q_y, q_{\dot{y}}, q_{I_{\max}}, q_{\sigma_{xy}})$. In 3D, the dynamical model is defined as $\mathbf{F}^1 = \mathrm{diag}(1,1,1,1,1,1,1,1,1)$ and the covariance matrix is $\mathbf{Q}^1 = \mathrm{diag}(q_x, q_{\dot{x}}, q_y, q_{\dot{y}}, q_z, q_{\dot{z}}, q_{I_{\max}}, q_{\sigma_{xy}}, q_{\sigma_z})$. For the directed motion model ($\omega = 2$), the dynamical model in 2D is defined as:

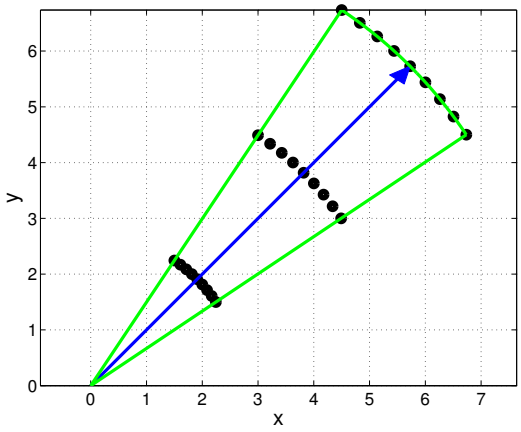
$$\mathbf{F}^2 = \begin{pmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.78}$$



(a) Low speed



(b) Medium speed



(c) High speed

Figure 3.7: Sample circular sector measurements (black dots). The predicted position is at the origin and the green lines delineate the circular sector. The extent of the circular sector depends on the magnitude of the velocity vector (speed) shown in blue. For each example $N_{\mathrm{cs},c} = 3$ and $N_{\mathrm{cs},j} = 8$, and so $N_{\mathrm{cs},c}(N_{\mathrm{cs},j} + 1) = 27$.

while the covariance matrix takes the following form:

$$\mathbf{Q}^2 = \begin{pmatrix} q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 & 0 \\ q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{I_{\max}} & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{\sigma_{xy}} \end{pmatrix}. \tag{3.79}$$

Likewise, in 3D the dynamical model is given by:

$$\mathbf{F}^2 = \begin{pmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.80}$$

with covariance matrix:

$$\mathbf{Q}^2 = \begin{pmatrix} q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{\mathrm{CV}}/3\Delta t^3 & q_{\mathrm{CV}}/2\Delta t^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{\mathrm{CV}}/2\Delta t^2 & q_{\mathrm{CV}}\Delta t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{I_{\max}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{\sigma_{xy}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{\sigma_z} \end{pmatrix}. \tag{3.81}$$

We have measurements $\mathbf{y} = (x', y', I'_{\max}, \sigma'_{xy})^T$ (2D case) or $\mathbf{y} = (x', y', z', I'_{\max}, \sigma'_{xy}, \sigma'_z)^T$

(3D case). For both models, the measurement matrix is given in 2D by:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.82}$$

The uncertainties of the measurements in 2D are represented by: $\mathbf{R} = \text{diag}(r_x, r_y, r_{I_{\max}}, r_{\sigma_{xy}})$. In 3D, the measurement matrix for both models is given by:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.83}$$

The uncertainties of the measurements in 3D are represented by: $\mathbf{R} = \text{diag}(r_x, r_y, r_z, r_{I_{\max}}, r_{\sigma_{xy}}, r_{\sigma_z})$.

Since the appearance model does not rely on the velocity components, the function $\Xi(\cdot;\cdot)$ is given by:

$$\Xi(\mathbf{y}; \hat{\mathbf{x}}) = \mathbf{H}^{+}\mathbf{y} \tag{3.84}$$

where $\mathbf{H}^{+} = \mathbf{H}^{T}$. The matrix $\mathbf{\Phi}$ used to project a measurement $\mathbf{y}$ onto a 2D or 3D position space are defined in (3.61) and (3.61), respectively. The transition matrix $\mathbf{\Delta}$ is given by:

$$\mathbf{\Delta} = \begin{pmatrix} 0.75 & 0.36 \\ 0.25 & 0.64 \end{pmatrix}. \tag{3.85}$$

The relatively higher values on the diagonal reflect that, for a single particle, there is a higher probability of remaining at a given motion model compared to the probability of changing to a different motion model (which is reflected by the off-diagonal values). The initial probabilities for the motion models are set to $P(\omega) = [\ 0.6\ \ 0.4\ ]^T$. We use the Lucas-Kanade optical flow approach [119] to initialize the velocity components.

3.9 Summary

This chapter presented our probabilistic tracking approaches. We discussed approaches based on the Kalman filter as well as particle filters. We also described our novel approach based on probabilistic data association and an ellipsoidal sampling scheme

(PDAE). To cope with multiple motion models, we also introduced an extension of the PDAE approach based on the IMM algorithm. We presented models for tracking round as well as elongated objects. In the next chapter, we describe an approach for behavior identification of fluorescent particles based on the trajectory information extracted by the developed tracking approaches.

Chapter 4

Identifying Virus-Cell Fusion in Multi-Channel Fluorescence Microscopy Image Sequences

Tracking yields the trajectories of fluorescent particles displayed in an image sequence. To obtain an explicit characterization of the behavior of the particles, object-based representations of the particles may be readily extracted from the trajectories. For example, the size over time of the particle may reveal certain behavior of the particle. Likewise, the intensity over time of fluorescent particles may reflect the current behavior of the particle. In this chapter, we introduce an automatic approach grounded within the theory of Bayesian estimation for identifying fusion of virus particles with the cell membrane based on the intensity over time of individual particles extracted from multi-channel fluorescence microscopy images [69]. Our approach adopts a *layered* architecture that decomposes the actions of a single virus particle into three primitives (viz., the intensity, the underlying temporal intensity model, and the behavior). The three primitives are represented by the different layers, which are in turn described using a stochastic hybrid system and a hidden Markov model (HMM). The two models are combined via a *maxbelief* strategy (e.g., [137]). To estimate the primitives we use a *hybrid particle filter* (e.g., [124]) as well as the Viterbi algorithm. Our layered probabilistic approach entails several advantages: First, the combination of stochastic hybrid systems and HMMs offer an improved modeling capability. Second, the maxbelief strategy is straightforward and introduces no additional computational overhead. Third, the modularity of such a layered approach endows our scheme with efficiency, since the dimensionality of the state space of the different layers is low. Because of the low dimensionality, a relatively low number of samples for the hybrid particle filter is sufficient to ensure a good support of the hybrid posterior. Additionally, the layers can also be adjusted; thus the approach is

flexible and can be straightforwardly adapted for identifying other behaviors. We also introduce temporal intensity models to describe the local fluctuations of the intensity of single virus particles. The models are defined as auto-regressive (AR) processes and thus allow an intuitive interpretation of the corresponding process parameters. To the best of our knowledge, this is the first time that a layered probabilistic approach has been introduced for identifying behaviors of fluorescent particles in multi-channel microscopy image sequences.

4.1 Overview of the Approach

Our layered probabilistic approach for behavior identification in multi-channel microscopy image data consists of three layers: The first layer corresponds to the intensity of a particle, and the second layer represents the temporal intensity model of the particle. The third and topmost layer models the behavior (including fusion) of the particle. The intensity is described via autoregressive (AR) processes while the temporal intensity models follow a first-order Markov chain. The two layers are jointly modeled using a stochastic hybrid system. We have applied the developed approach to two-channel image sequences displaying virus particles. The hybrid stochastic system describes the intensity and temporal intensity models of one channel and is thus used separately on the temporal intensity statistics of each channel of the two-channel image data. To take into account the information from both channels, the identified temporal intensity models in the individual channels are fed to the layer which models the behavior. The behavior is described via an HMM. By using the Viterbi algorithm we obtain the most probable sequence of behaviors for a particle.

4.2 Stochastic Hybrid Systems of Virus Intensity: Bayesian Framework

In our approach for identifying fusion of virus particles with the cell membrane we represent the intensity of a virus particle via a set of temporal intensity models that take the form of non-linear autoregressive processes. At a given time step it is assumed that only one temporal intensity model specifies the observed intensity of the particle. The aim is to estimate the intensity as well as the current temporal intensity model from a given sequence of observed intensities. We formulate this task as a sequential hybrid estimation problem. Within the theory of stochastic hybrid systems (e.g., [82], [57]), it is assumed that a variable of interest of a system can be characterized via a set A of N_A predefined *system models*. At time step t, the variable is both governed by a

certain model $\alpha_t \in A$, and manifested via a state vector $\mathbf{x}_t$ (in our case, the intensity of a particle), which in turn is reflected by a noisy measurement $\mathbf{y}_t$ (in our case, the intensity of a particle obtained from the image data). The goal is to estimate $(\mathbf{x}_t,\ \alpha_t)$ given a sequence of measurements $\mathbf{y}_{1:t}$. A Bayesian approach involves computing the posterior densities $p(\mathbf{x}_t, \alpha_t|\mathbf{y}_{1:t})$, which can be factored as:

$$p(\mathbf{x}_t, \alpha_t|\mathbf{y}_{1:t}) = p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t})P(\alpha_t|\mathbf{y}_{1:t}). \tag{4.1}$$

This implicates computing the model-conditioned posterior densities $p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t})$ as well as the posterior model probabilities $P(\alpha_t|\mathbf{y}_{1:t})$. At time step $t-1$, all N_A model-conditioned posteriors are maintained. Since the evolution of the model α_{t-1} is represented using a Markov chain associated with a transition matrix $\mathbf{\Pi}$, each model $\alpha_{t-1} = j$ may 'branch' into any other model $\alpha_t = i$ at the next time step t with probability π_{ij}. The optimal solution involves computing each possible *model history* ([193], [82]), i.e., computing the full model history tree, which leads to an exponential increase in computational effort w.r.t. time. Suboptimal strategies aim at maintaining a constant computational load over time. One such strategy (e.g., [22], [57], [124]) with application to biological particles (e.g., [64], [178]) involves first computing the prior model probabilities $P(\alpha_t|\mathbf{y}_{1:t-1})$ via the Chapman-Kolmogorov equation for the Markov chain underlying the model evolution:

$$P(\alpha_t|\mathbf{y}_{1:t-1}) = \sum_{\alpha_{t-1}=1}^{N_A} P(\alpha_{t-1}|\mathbf{y}_{1:t-1})P(\alpha_t|\alpha_{t-1}), \tag{4.2}$$

The branches entailed by the evolution of the Markov chain are 'mixed' into N_A model-conditioned prior densities $p(\mathbf{x}_{t-1}|\alpha_t, \mathbf{y}_{1:t-1})$. Assuming that $\mathbf{x}_{t-1}$ is independent of α_t given α_{t-1}, the mixing of the densities amounts to a weighted sum of the model-conditioned posteriors:

$$p(\mathbf{x}_{t-1}|\alpha_t, \mathbf{y}_{1:t-1}) = \sum_{\alpha_{t-1}=1}^{N_A} p(\mathbf{x}_{t-1}|\alpha_{t-1}, \mathbf{y}_{1:t-1})P(\alpha_{t-1}|\alpha_t, \mathbf{y}_{1:t-1}). \tag{4.3}$$

Here the weights are obtained by applying Bayes' theorem:

$$P(\alpha_{t-1}|\alpha_t, \mathbf{y}_{1:t-1}) = \frac{P(\alpha_{t-1}|\mathbf{y}_{1:t-1})P(\alpha_t|\alpha_{t-1})}{P(\alpha_t|\mathbf{y}_{1:t-1})}. \tag{4.4}$$

Each of these N_A mixture prior densities can now be employed within a model-matched filtering framework. Concretely, the model-conditioned mixture priors in (4.3) are propagated over time by applying the corresponding model-conditioned *dynamical model*

$p(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_t)$:

$$p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_t)\, p(\mathbf{x}_{t-1}|\alpha_t, \mathbf{y}_{1:t-1})\, d\mathbf{x}_{t-1}. \tag{4.5}$$

Using a model-conditioned *measurement model* $p(\mathbf{y}_t|\mathbf{x}_t, \alpha_t)$, an update of the predicted densities $p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t-1})$ is obtained via Bayes' theorem:

$$p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t, \alpha_t)\, p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t-1}). \tag{4.6}$$

Likewise, the predicted probabilities $P(\alpha_t|\mathbf{y}_{1:t-1})$ are updated using Bayes' theorem:

$$P(\alpha_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\alpha_t, \mathbf{y}_{1:t-1})\, P(\alpha_t|\mathbf{y}_{1:t-1})}{\sum\limits_{\alpha_t} p(\mathbf{y}_t|\alpha_t, \mathbf{y}_{1:t-1})\, P(\alpha_t|\mathbf{y}_{1:t-1})}, \tag{4.7}$$

where the denominator acts as a normalization factor, and where, by the law of total probability and assuming that the measurements are independent of each other, the likelihood $p(\mathbf{y}_t|\alpha_t, \mathbf{y}_{1:t-1})$ is defined as:

$$p(\mathbf{y}_t|\alpha_t, \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_t, \alpha_t) p(\mathbf{x}_t|\alpha_t, \mathbf{y}_{1:t-1}) d\mathbf{x}_t. \tag{4.8}$$

This strategy consists of $N_A + 1$ estimation processes: N_A model-matched filters (see (4.5) and (4.6)) as well as an estimator for the model (see (4.2) and (4.7)). Note that the N_A model-matched filters interact with the model estimator via (4.3) as well as via (4.7). However, handling the N_A mixture priors in (4.3) that are used as input for the model-matched filters could be cumbersome, since these priors can have a complex form. Algorithms adopting this strategy therefore reduce the complexity of the priors by pruning or merging the components of the involved mixture densities. For example, the Interacting Multiple Model (IMM) algorithm ([21], [22]; see Section 3.8.3) assumes that the models are linear and Gaussian and thus each prior amounts to a Gaussian mixture. The IMM converts this Gaussian mixture to a single Gaussian density of matched first and second moments (see (3.51) and (3.52)) that is amenable to analytical techniques based on the Kalman filter. In cases where the models are non-linear and/or non-Gaussian, exact calculations of the above (multivariate) relations are in general not possible. However, one may obtain a sound numerical approximation of these distributions via particle filters. An approach combining the IMM algorithm with particle filters has been introduced in [57] (see also [178]). Whereas in the standard IMM each prior (4.3) is approximated by a single Gaussian density, in [57] each prior is more accurately represented by a non-parametric mixture density with N_A components, where

each component is approximated by a particle filter with a fixed number of samples. In comparison to that, in our work we resolve the recursions using a hybrid particle filter that approximates the priors by directly simulating the dynamical models governing the evolution of the hybrid system over time (i.e., each sample represents an instance of the hybrid system at a given time step, cf. [124]). We use the hybrid particle filter to estimate both the discrete and continuous variables, and we do not assume a Gaussian density for the likelihood of each model. We use N_s samples for all N_A models, and the number of samples for each model varies over time. Using a large number of samples N_s relative to the dimension of the state space allows our approach to cope well with relatively small values for the transition probabilities, which represent rare events.

4.3 Stochastic Hybrid Systems of Virus Intensity: Implementation via a Hybrid Particle Filter

Based on the principle of *importance sampling*, a set $\{\mathbf{x}^k; w^k\}_{k=1}^{N_s}$ of N_s random samples $\mathbf{x}^k$ (the 'particles') associated with importance weights w^k is assumed to approximately represent an arbitrary distribution $p(\mathbf{x})$ in the sense that selecting a certain $\mathbf{x}^k$ with probability proportional to w^k amounts to drawing a random sample from $p(\mathbf{x})$. The particle filter (e.g., [72], [88]; see Section 3.6) is a numerical scheme for recursively constructing such a set $\{\mathbf{x}_t^k; w_t^k\}_{k=1}^{N_s}$ for approximating the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, which is defined by stochastic propagation and Bayes' theorem as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\, d\mathbf{x}_{t-1}. \tag{4.9}$$

The weighted approximation of the posterior obtained by the particle filter is thus given as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{k=1}^{N_s} w_t^k\, \delta(\mathbf{x}_t - \mathbf{x}_t^k), \tag{4.10}$$

where $\delta(\cdot)$ denotes the Dirac delta measure and $\sum_{k=1}^{N_s} w_t^k = 1$. Here we recall the main steps of the particle filter. More details are presented in Section 3.6. Briefly, the particle filter resolves the (single model) Bayesian recursion defined in (4.9) in three steps. Starting with the set $\{\mathbf{x}_{t-1}^k; w_{t-1}^k\}_{k=1}^{N_s}$, the filter first generates each predicted sample $\mathbf{x}_t^k$ with a random draw from the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1}^k)$. Second, these predicted samples are re-weighted according to $w_t^k \propto p(\mathbf{y}_t|\mathbf{x}_t^k)$. The weights are normalized so that they sum up to unity. Finally, to select the most representative samples, the filter resamples N_s times from the set of weighted samples $\{\mathbf{x}_t^k; w_t^k\}_{k=1}^{N_s}$.

One approach for extending the particle filter to stochastic hybrid systems consists

of augmenting each k-th sample with a model index α_t^k thereby obtaining a set of *hybrid* samples $\{\mathbf{x}_t^k, \alpha_t^k; w_t^k\}_{k=1}^{N_s}$ (e.g., [124]). Based on this set of hybrid samples, the posterior in (4.1) can be represented by the following weighted approximation ([8]):

$$p(\mathbf{x}_t, \alpha_t|\mathbf{y}_{1:t}) \approx \sum_{k=1}^{N_s} w_t^k \, \delta(\mathbf{x}_t - \mathbf{x}_t^k)\mathbf{1}_{\alpha_t^k}(\alpha_t). \tag{4.11}$$

Here $\mathbf{1}_{\cdot}(\cdot)$ is the indicator function, where $\mathbf{1}_{\alpha^k}(\alpha) = 1$ if $\alpha^k = \alpha$, else $\mathbf{1}_{\alpha^k}(\alpha) = 0$. At time step t the *hybrid particle filter* maintains N_A model-conditioned posterior densities $p(\mathbf{x}_t|\alpha_t = i, \mathbf{y}_{1:t})$, where $i \in A$. Each i-th model-conditioned posterior is thus approximated via the set of samples $\{\mathbf{x}_t^k, \alpha_t^k = i\}$:

$$p(\mathbf{x}_t|\alpha_t = i, \mathbf{y}_{1:t}) \approx \sum_{k \in \Upsilon_i} \tilde{w}_t^k \, \delta(\mathbf{x}_t - \mathbf{x}_t^k), \tag{4.12}$$

where $\Upsilon_i = \{k \,|\, \mathbf{x}_t^k, \alpha_t^k = i\}$ and $\tilde{w}_t^k = \frac{w_t^k}{\sum_{l \in \Upsilon} w_t^l}$. Similarly, the posterior model probabilities $P(\alpha_t|\mathbf{y}_{1:t})$ are approximated by the proportion of samples supporting model i:

$$P(\alpha_t = i|\mathbf{y}_{1:t}) \approx \frac{|\{\mathbf{x}_t^k, \alpha_t^k = i\}|}{N_s}, \tag{4.13}$$

where $|\cdot|$ denotes the set size operator. The hybrid particle filter recursively constructs a set of hybrid samples as follows. At time step $t-1$, the predicted model index α_t^k is generated by drawing a random sample from $P(\alpha_t|\alpha_{t-1}^k = j)$, which is defined by the j-th column of the transition matrix $\mathbf{\Pi}$. Samples that branch from model $\alpha_{t-1}^k = j$ into model $\alpha_t^k = i$ represent approximately the branched priors $p(\mathbf{x}_{t-1}|\alpha_{t-1} = j, \alpha_t = i, \mathbf{y}_{1:t-1})$. Following the mixing strategy as expressed by (4.3), the model-conditioned prior densities $p(\mathbf{x}_{t-1}|\alpha_t = i, \mathbf{y}_{1:t-1})$ are supported by the set of samples with predicted model index $\alpha_t^k = i$. The particle-based representation of each i-th model-conditioned prior is suitable for the sequential importance sampling-resampling scheme of the standard particle filter. Specifically, a prediction for $\mathbf{x}_t^k$ is obtained by drawing a random sample from the dynamical model $p(\mathbf{x}_t|\mathbf{x}_{t-1}^k, \alpha_t^k = i)$ conditioned on the predicted model $\alpha_t^k = i$. The predicted samples are assumed to support the predicted density $p(\mathbf{x}_t|\alpha_t = i, \mathbf{y}_{1:t-1})$. Next, the weights of the predicted samples are re-calculated according to the model-conditioned measurement model: $w_t^k \propto p(\mathbf{y}_t|\mathbf{x}_t^k, \alpha_t^k = i)$. The resampling step is finally applied to the entire set of samples to obtain a set $\{\mathbf{x}_t^k, \alpha_t^k; w_t^k\}_{k=1}^{N_s}$ of N_s hybrid samples. At the end of each iteration, a maximum a posteriori (MAP) estimate for the model α_t may be obtained from the model posterior probabilities (e.g., [87]):

$$\hat{\alpha}_t = \arg\max_i P(\alpha_t = i|\mathbf{y}_{1:t}). \tag{4.14}$$

An estimate for $\mathbf{x}_t$ may be computed as the weighted mean over all N_s samples:

$$\bar{\mathbf{x}}_t = \sum_{k=1}^{N_s} w_t^k \, \mathbf{x}_t^k. \tag{4.15}$$

Alternatively, one may obtain the mean estimate of the model-conditioned posterior corresponding to the MAP estimate $\hat{\alpha}_t$:

$$\bar{\mathbf{x}}_t = \sum_{k \in \hat{\Upsilon}} \tilde{w}_t^k \, \mathbf{x}_t^k, \tag{4.16}$$

where $\hat{\Upsilon} = \{k \,|\, \mathbf{x}_t^k, \alpha_t^k = \hat{\alpha}_t\}$ and $\tilde{w}_t^k = \frac{w_t^k}{\sum_{l \in \hat{\Upsilon}} w_t^l}$.

4.4 Hidden Markov Model of Virus Behavior

In our application, we study pseudotyped HIV-1 particles where fusion with the cell membrane in two-channel microscopy image sequences is characterized by *both* the loss of the fluorescent label attached to the outer shell in one channel as well as the preservation of the label attached to the particle's inner part in the other channel. Thus, the intensity as well as the underlying temporal intensity models of the two channels must be considered when automatically identifying individual fusion events. Using stochastic hybrid systems for identifying the temporal intensity models of both channels entails increasing the dimensionality of the state vector $\mathbf{x}_t$ as well as increasing the number of models N_A. In the worst case, one would need to define N_A^2 models accounting for the *joint* dynamics of the intensities of both channels. This would require building a transition matrix $\mathbf{\Pi}$ of size $N_A^2 \times N_A^2$, which is generally impractical to define. Because of the scarce amount of available training data, learning the transition matrix may lead to inaccurate estimates of the transition probabilities. Additionally, increasing the number of models N_A requires a significant increase in the number of samples for the hybrid particle filter to obtain a comparable performance. Instead of using this monolithic strategy, we adopt the following hierarchical solution. First, we use hybrid particle filters for the intensities measured on each channel separately. This yields a sequence $\hat{\alpha}_{1:T}^c$ of T temporal intensity models per channel c that correspond to the MAP model estimates, i.e., the models with the highest *belief* obtained at each time step t. The sequence of vectors $\boldsymbol{\alpha}_{1:T}$, which comprises the inferred models of both channels, is fed to a Hidden Markov Model (HMM) (e.g., [147]) that maps these sequences to a sequence of behaviors $\beta_{1:T}$, where β represents a certain behavior from a set B of N_B predefined behaviors. Concretely, a Markov chain with transition matrix $\mathbf{\Phi}$ is assumed to underlie the sequence of behaviors $\beta_{1:T}$. Using the HMM formalism, one can define a posterior

distribution over the sequence of behaviors $\beta_{1:T}$ given a sequence of observed intensity models $\boldsymbol{\alpha}_{1:T}$ as follows:

$$P(\beta_{1:T}|\boldsymbol{\alpha}_{1:T}) \propto \prod_{t=1}^{T} P(\beta_t|\beta_{t-1})\, P(\boldsymbol{\alpha}_t|\beta_t), \tag{4.17}$$

where $P(\beta_t = i|\beta_{t-1} = j)$ is given by the corresponding element ϕ_{ij} of the transition matrix $\boldsymbol{\Phi}$. $P(\boldsymbol{\alpha}_t|\beta_t)$ denotes the measurement model defined by the HMM. Our aim is to determine the MAP sequence of behaviors:

$$\hat{\beta}_{1:T} = \arg\max_{\beta_{1:T}} P(\beta_{1:T}|\boldsymbol{\alpha}_{1:T}). \tag{4.18}$$

Note that (4.18) can be efficiently solved using the Viterbi algorithm (e.g., [62]). Thus, our approach represents the behavior β, the temporal intensity model α, as well as the intensity statistics $\mathbf{x}$ via three stacked Markov processes that amount to a layered probabilistic approach where a *maxbelief* strategy ([137]) is used for linking the two top layers (namely behaviors and intensity models) in a bottom-up fashion.

4.5 Model Definitions

4.5.1 Temporal Intensity Models

In this section, we define explicit representations of the temporal intensity models underlying the observed intensities. The models must account for the various phenomena exhibited by the observed intensity values. For example, because of photobleaching, the intensity exhibits a slow downward trend. In comparison, fusion with the cell membrane is characterized by a steep decrease in the intensity of one channel. The models should account for the different magnitudes of the change in intensity. In our case, the state vector $\mathbf{x}$ is given by the scalar value corresponding to the intensity I (which refers to the object intensity plus the background intensity). We define $N_A = 3$ intensity models: *constant intensity* (CI, $\alpha = 1$), *positive intensity change* (PIC, $\alpha = 2$), and *negative intensity change* (NIC, $\alpha = 3$). The corresponding dynamical models $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \alpha_t)$ are described via first-order auto-regressive processes on the state vector $\mathbf{x}_t$, which take the following form:

$$\mathbf{x}_t = f_{\alpha_t}(\mathbf{x}_{t-1}, v_{\alpha_t}), \tag{4.19}$$

where v_{α_t} is a noise variable sampled from a zero-mean Gaussian distribution with variance Q_{α_t}. To define the function $f(\cdot,\cdot)$ in (4.19), the CI model ($\alpha = 1$) adopts the

following relation:

$$f_{\alpha_t}(\mathbf{x}_{t-1}, v_{\alpha_t}) = I_{t-1} + v_{\alpha_t}. \tag{4.20}$$

For this model, we define $Q_{\alpha_t} = q_{\text{CI}}$, where q_{CI} is the variance of the deviation in the intensity over a time interval. For the PIC model ($\alpha = 2$), the function $f(\cdot, \cdot)$ in (4.19) is given as:

$$f_{\alpha_t}(\mathbf{x}_{t-1}, v_{\alpha_t}) = I_{t-1} + |v_{\alpha_t}| \tag{4.21}$$

and $Q_{\alpha_t} = q_{\text{PIC}}$. Likewise, the NIC model ($\alpha = 3$) defines $f(\cdot, \cdot)$ in (4.19) as:

$$f_{\alpha_t}(\mathbf{x}_{t-1}, v_{\alpha_t}) = I_{t-1} - |v_{\alpha_t}| \tag{4.22}$$

and $Q_{\alpha_t} = q_{\text{NIC}}$. Analogous to the CI model, q_{PIC} and q_{NIC} regulate the deviation for the intensity variable over time. With a relatively small perturbation of the intensity, as regulated by q_{CI}, the CI model captures smooth changes in the intensity introduced by photobleaching as well as other auto-fluorescent structures. By using a relatively large value for q_{NIC}, the NIC model represents sharp decreases in the intensity, which in our application correspond to the specific behavior of fusion. The PIC model is added for completeness. In addition to the model-conditioned dynamical model, each temporal intensity model α could in principle define a model-conditioned measurement model $p(\mathbf{y}_t|\mathbf{x}_t, \alpha_t)$. In our case all models share a common measurement model, i.e., $p(\mathbf{y}_t|\mathbf{x}_t, \alpha_t) = p(\mathbf{y}_t|\mathbf{x}_t)$. The proposed measurement model quantifies the probability that the predicted intensity I matches the particle's mean intensity y measured from the image:

$$p(y|\mathbf{x}) \propto \exp\left(-\frac{D(y, I)^2}{2\sigma_n^2}\right), \tag{4.23}$$

where $D(\cdot)$ is the Euclidean distance and σ_n describes the expected level of noise.

In our application, virus particles typically exhibit a constant behavior and fusion is rare. In other words, we assume that once the intensity follows a certain temporal model, the intensity will not jump between models but instead will adhere to a model for a certain time period. The entries π_{ij} of the transition matrix $\mathbf{\Pi}$ reflect these considerations. Because of this model steadiness, the matrix $\mathbf{\Pi}$ is strongly diagonal, i.e., the entries on the diagonal have larger values than the off-diagonal entries. For stability reasons, we also assume that the intensity cannot switch directly from a PIC model to a NIC model, or vice-versa. While the transition matrix could be learned from training data, we manually set the entries for the following reasons. First, large amounts of training data would be required to obtain sound estimates for the transition probabilities, especially since fusion of HIV-1 particles with the cell membrane is relatively rare ([126]), and thus the number of exemplary periods of intensity changes (e.g., decrease of inten-

sity) is inherently low. In addition, this training data would need to include one model label for each time step. Such a labeling would be quite tedious and time-consuming to obtain. Second, manually setting the values allows us to intuitively incorporate prior knowledge on the transition between the models. This also gives us a fine control over the approach. Certainly there is some risk that a user may adjust the values until intended results are obtained. However, in our case this risk is relatively small since the values for the transition probabilities were kept fixed for all our experiments.

4.5.2 Model of Virus Behavior

The behavior of HIV-1 particles throughout the cell entry process is not fully understood and thus this topic is currently the subject of intensive research ([197]). Initially, HIV-1 was assumed to enter the cell exclusively by fusion with the cell membrane. This has been recently challenged by [126], who contend that endocytosis, which is an alternative entry mechanism, also plays a role in the cell entry process. Additionally, the process is dependent on a large number of factors (e.g., sequence of the viral envelope protein, or the presence and density of receptor and co-receptors on the cell surface) whose description would lead to rather intricate models that still need to be established. Here, we restrict ourselves to abstractions of the behavior relative to the inferred temporal intensity models. A straightforward definition for these general behaviors would involve only two abstractions: fusion and non-fusion. We refrain from this approach as it leads to a coarse mapping between the behaviors and the intensity models. While certainly fusion (along with the corresponding temporal intensity models) is the main behavior that we aim to identify, other combinations of temporal intensity models correspond to behaviors that could be of interest to biologists. In our case, we define $N_B = 4$ general behaviors. One behavior is introduced for instances where the intensity corresponding to the outer shell's label remains constant or increases; we denote this as the *Outer Shell Constant or Increasing* (OSCI) behavior ($\beta = 1$). The preservation of the label attached to the particle's inner part is a prerequisite for fusion, and thus we assign another behavior to these types of situations, which are referred as the *Inner Part Preservation* (IPP) behavior ($\beta = 2$). The loss of the fluorescent label attached to the outer shell is a precondition for fusion too, but also an interesting action in itself, and thus we specify one general behavior to such occurrences; we call this the *Outer Shell Loss* (OSL) behavior ($\beta = 3$). Finally, both the preservation of the inner part's label as well as the loss of the outer shell's label are indicative of *fusion* (FUSION, $\beta = 4$). The transition probabilities for this 4-state Markov chain could be learned from training data. However, since we are working with high level abstractions of the behavior, we choose manually the transition probabilities for this 4-state Markov chain. This has worked

well in practice. For instance, since in our experimental setup the duration of the fusion behavior is relatively short, we have set the probability for the transitions within the fusion behavior to $\phi_{33} = 0.5$. This mounts to an average duration of two frames, which turned out to work well as a prior. The probabilities for transitions from and to the fusion behavior are distributed almost equally. Similar considerations apply to the other behaviors. This yields a flexible model which does not strongly favor transitions towards a particular behavior, where the behaviors have similar durations, and where transitions between all behaviors are allowed, i.e., $\phi_{ij} \neq 0 \ , \forall \, i, j \in B$. Note that the entries on the diagonal should have higher values than the off-diagonal entries. This reflects the fact that there is usually a higher probability for the particle to remain within the same behavior than to jump to another behavior. Nonetheless the off-diagonal values should be relatively large as this would allow a relatively fast change between behaviors.

To define the measurement model $P(\boldsymbol{\alpha}_t|\beta_t)$ for the HMM, we map the behavior β_t as well as the vector $\boldsymbol{\alpha}_t = (\hat{\alpha}_t^0, \hat{\alpha}_t^1)$ of the temporal intensity model estimates to a common feature space. The chosen feature space should preserve a certain order for the behaviors, since some behaviors are more closely related (e.g., the FUSION behavior is related to the OSL behavior but it is more distant from the OSCI behavior). To preserve this intuitive ordering over the different behaviors, we map the behaviors to a two-dimensional Hamming space (e.g., [12]), which is a metric space defined by the two-fold Cartesian product of the set $\{0, 1\}$, along with the Hamming distance $D_H(\cdot, \cdot)$. The mapping is achieved via the function $h : B \rightarrow \{0, 1\}^2$, where $h(\beta)$ is the binary representation of the behavior indices and is defined as follows: $h(0) = (0, 0)$, $h(1) = (0, 1)$, $h(2) = (1, 0)$, and $h(3) = (1, 1)$. Similarly, we map the vector $\boldsymbol{\alpha}$ to this Hamming space using the surjective function $g : A^2 \rightarrow \{0, 1\}^2$, where $g(\boldsymbol{\alpha})$ is given in Table 4.1. Note that by adjusting the function $g(\cdot)$ we can cope with different experimental conditions (e.g., in some cases the simultaneous loss of both fluorescent labels may indicate fusion). Our definition of $g(\cdot)$ also implies that channel $c = 0$ corresponds to the intensity of the fluorescent label attached to the outer shell, while channel $c = 1$ represents the intensity of the label tagged to the inner part of the virus particle. To quantify the similarity between $\boldsymbol{\alpha}$ and β via $P(\boldsymbol{\alpha}|\beta)$, we use the following exponential function:

$$P(\boldsymbol{\alpha}|\beta) \propto \exp\left(-\frac{D_H(g(\boldsymbol{\alpha}), h(\beta))^2}{2\sigma_H^2}\right), \tag{4.24}$$

where σ_H inversely regulates the level of fidelity in the Hamming space (we used $\sigma_H = 1$).

Table 4.1: Function $g(\boldsymbol{\alpha})$.

$\boldsymbol{\alpha}$	$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$	$(2,0)$	$(2,1)$	$(2,2)$
$g(\boldsymbol{\alpha})$	$(0,1)$	$(0,0)$	$(0,0)$	$(0,1)$	$(0,0)$	$(0,0)$	$(1,1)$	$(1,0)$	$(1,0)$

4.6 Summary

In this chapter, we presented a probabilistic approach for identifying behaviors of fluorescent particles. In particular, we developed an approach for identifying fusion between a virus particle and a cell based on the intensity over time of an individual virus particle. The proposed approach is grounded within a Bayesian framework and follows a layered architecture. Note that our approach for behavior identification builds upon a virus tracking approach (see Chapter 3) which is performed prior to applying our approach for behavior identification. In the next chapter, we evaluate experimentally the performance of both the tracking approaches as well as the behavior identification approach.

Chapter 5

Experimental Results

In this chapter, we present our experimental results [70], [69] for the developed tracking approaches (Chapter 3) as well as for the developed behavior identification approach (Chapter 4). To study the performance of the approaches, we use synthetic data as well as real microscopy image sequences displaying HIV-1 particles. We defined performance measures and quantify the performance of the approaches with respect to ground truth. For the real images, the ground truth is determined manually. We have also performed a quantitative comparison of the performance of the approaches with previous approaches. In addition, we present tracking results obtained for image sequences displaying HSV compartments as well as for image sequences displaying microtubule tips within a neuron's growth cone.

5.1 Tracking: Experimental Results

We evaluate the performance of several deterministic and probabilistic tracking approaches, including approaches based on the Kalman filter and the particle filter. We also perform a detailed evaluation of our new PDAE tracking approach.

5.1.1 Performance Evaluation of Deterministic and Probabilistic Approaches Based on Real 2D Image Sequences

We perform an extensive evaluation of eight different deterministic and probabilistic tracking approaches based on nine real microscopy image sequences displaying HIV-1 particles [70]. Such an evaluation is essential to determine the performance of the approaches under realistic conditions. Below, we describe the evaluated approaches and present the results.

5.1.1.1 Evaluated Tracking Approaches

We evaluate eight different tracking approaches. All approaches employ a 2D Gaussian intensity model to represent fluorescent particles. The deterministic approaches are defined based on the different combinations of the localization schemes (Section 3.1) and the motion correspondence scheme (Section 3.3). We evaluate two deterministic approaches: 1) spot-enhancing filter with global nearest neighbor (SEF&GNN), and 2) 2D Gaussian fitting with global nearest neighbor (GaussFit&GNN). Note that the motion correspondence step requires a set $\hat{Y}$ of predicted measurements. For the deterministic schemes, a "predicted" measurement $\hat{\mathbf{y}}_t^i$ is given by the latest position estimate obtained for the tracked fluorescent particle i at the previous time step $t-1$.

The probabilistic approaches are analogously obtained by different combinations of the localization schemes and the spatial-temporal filters. The filters use the models for tracking round particles with a random walk motion model described in Section 3.8.1. We investigate the following probabilistic approaches: 3) spot-enhancing filter and Kalman filters (Section 3.5) (SEF&Kalman), 4) spot-enhancing filter and a mixture of particle filters (e.g., [203], [176]) (SEF&MPF), 5) spot-enhancing filter and independent particle filters (Section 3.6) (SEF&IPF), 6) 2D Gaussian fitting and Kalman filters (GaussFit&Kalman), 7) 2D Gaussian fitting and a mixture of particle filters (GaussFit&MPF), and 8) 2D Gaussian fitting and independent particle filters (GaussFit&IPF). Note that the approaches using a mixture of particle filters can only track a fixed number of objects. The approaches using Kalman filters as well as those using independent particle filters require a motion correspondence step. For these approaches, we have employed the global nearest neighbor approach described in Section 3.3.

5.1.1.2 Performance Measure

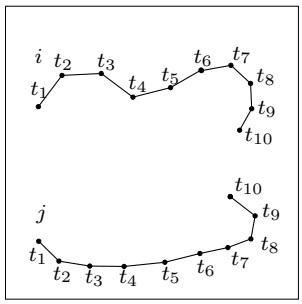
To quantitatively evaluate the performance of the tracking approaches, we compute the tracking accuracy P_{track}:

$$P_{\text{track}} = \frac{n_{\text{track,correct}}}{n_{\text{track,total}}}, \tag{5.1}$$

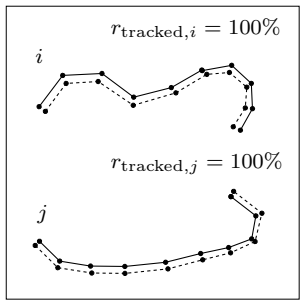
where $n_{\text{track,correct}}$ is the number of correctly computed trajectories and $n_{\text{track,total}}$ represents the number of true trajectories. The value $n_{\text{track,correct}}$ is computed as the weighted sum of the percentage of tracked time steps $r_{\text{tracked},i}$ for each *i-th* true trajectory:

$$n_{\text{track,correct}} = \sum_{i=1}^{n_{\text{track,total}}} w_i r_{\text{tracked},i}, \tag{5.2}$$

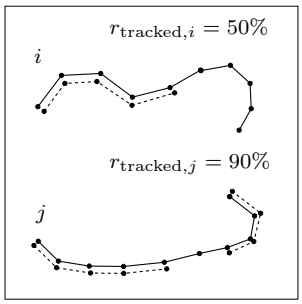
where the weight w_i is given by a Gaussian weighting scheme, which takes as its argument the number of correctly computed trajectories $n_{\text{track},i}$ corresponding to each *i-th* true



(a) True trajectories for two objects i and j represented by solid lines. Each trajectory comprises 10 time steps; $n_{\text{track,total}} = 2$.



(b) True trajectories (solid lines) and trajectories computed by an ideal tracking approach (dashed lines). In this case, $n_{\text{track,correct}} = 2$, and thus $P_{\text{track}} = 100\%$.



(c) True trajectories (solid lines) and trajectories computed by a non-ideal tracking approach (dashed lines). In this case, $n_{\text{track,correct}} = 1.04$, and thus $P_{\text{track}} = 52\%$.

Figure 5.1: Examples illustrating the performance measure P_{track}.

Table 5.1: Description of real image sequences.

	Dimensions [pixels]	No. of time steps	No. of objects
Seq. 1	256×256	250	23
Seq. 2	256×256	250	10
Seq. 3	256×256	250	5
Seq. 4	256×256	150	21
Seq. 5	512×512	200	15
Seq. 6	512×512	400	29
Seq. 7	512×512	400	31
Seq. 8	512×512	400	43
Seq. 9	512×512	400	24

trajectory: $w_i = \mathcal{N}(n_{\text{track},i}; \mu = 1, \sigma = 1)$. Since ideally one expects a single correctly computed trajectory for each *i-th* true trajectory (i.e., ideally $n_{\text{track},i} = 1$), we set the mean of the Gaussian weighting scheme to $\mu = 1$. The weighting scheme is introduced to penalize computed trajectories that are broken. A computed trajectory is assumed to be correct if the error (Euclidean distance) between the measured object position and the true object position is below 2 pixels. Examples illustrating the measure P_{track} are shown in Figure 5.1.

5.1.1.3 Experimental Setup

For our evaluation we used real microscopy image sequences displaying double fluorescently labeled HIV-1 particles. The particles were imaged using a fluorescence wide-field setup based on a Zeiss Axiovert 200 M microscope with a Roper Scientific Cascade II EM-CCD camera. Fluorophores were excited with their respective excitation wave-

lengths and a pair of images (red and green images) was acquired every 100 ms. The spatial resolution is 160 nm/pixel ([108]). Note that the number of virus particles, the level of cellular autofluorescence, the image dimensions, the number of images, and the level of image noise vary largely for the evaluated image sequences (see Figure 1.1 and Table 5.1). Ground truth for the virus positions was obtained by manual tracking using the commercial software MetaMorph. For all sequences, we use fixed parameter values for all approaches. For the approaches using the spot-enhancing filter, a factor of $c = 3.7$ in (3.1) is used. For the approaches using the global nearest neighbor scheme (cf. Section 3.3), the displacement parameter is set to $d_{max} = 2$ pixels and the disappearance duration is set to 1 frame i.e., a particle can disappear for at most one frame. For the approaches using independent particle filters (IPF), we use $N_s = 1000$ samples per object. Similarly, for the approaches using a mixture of particle filters (MPF), each mixture component is initially assigned 1000 samples. The noise parameters for the dynamical model of the Kalman filter are set analogously as the ones employed for the particle filter.

5.1.1.4 Results

Let us first consider one of the evaluated image sequences ("Seq. 4") consisting of 150 images (256×256 pixels, 16-bit). To assess the performance of the algorithms in areas with a relatively high object density, we select a region-of-interest (ROI) of 66×66 pixels. We first discuss the results for this ROI; the results for the whole sequence are discussed thereafter. In this sequence, some of the virus particles go out of focus, thereby introducing blurring effects. This in turn leads to difficulties for the virus detection schemes, since they rely on the apparent size and intensity of the particles. Photobleaching effects are also noticeable. Sample original images and results obtained with the approaches using 2D Gaussian fitting are shown in Figure 5.2. It turned out that the approaches based on particle filters can track particles with a decreasing contrast. For instance, consider the particle at the lower right corner of the ROI (see the arrows). The intensity contrast as well as the size of this particle decreases during the initial time steps. The particle localization algorithms eventually fail to detect this particle. For the deterministic approaches as well as for the Kalman filter, detection failures lead to broken trajectories or no trajectories are obtained. The particle filter, in contrast, manages such situations by virtue of its probabilistic 'recognition-by-synthesis' approach. That is, given the previously estimated appearance parameters, the particle filter is able to generate variations of the particle's appearance, which are directly evaluated against the image data by means of its non-linear measurement model. By generating samples that resemble the distorted appearance of the particle, the particle filter is able to identify

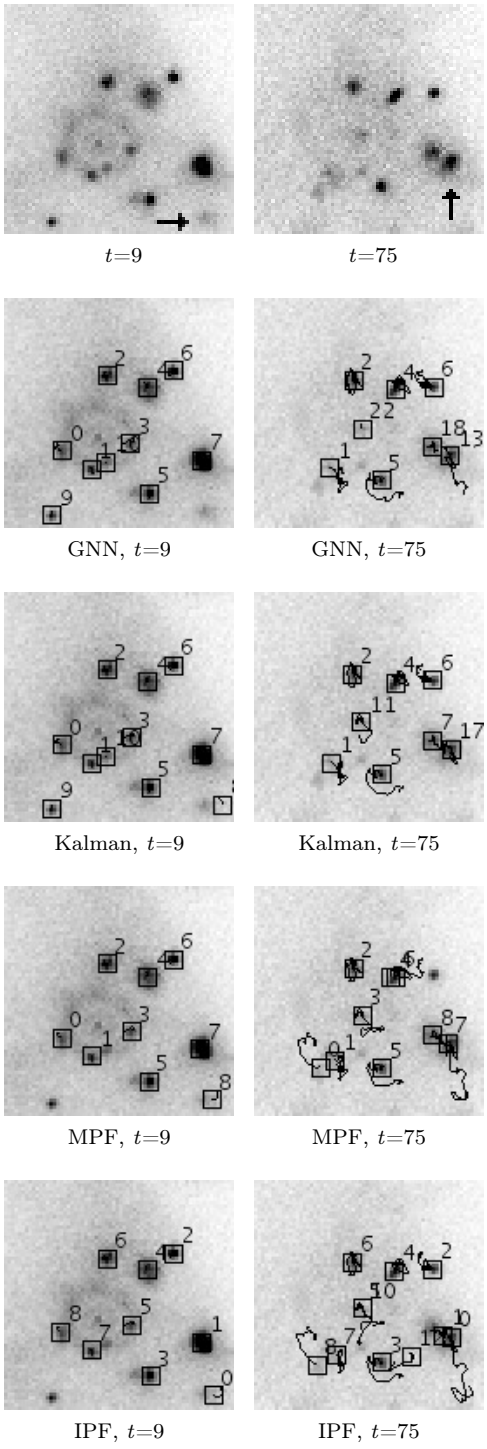


Figure 5.2: Original images (section, first row) and tracking results (second to fifth row) for the real image sequence "Seq. 4". For all shown results 2D Gaussian fitting is used for particle localization. (Image intensities have been inverted for visualization purposes).

such a particle.

These remarks are supported by the tracking accuracy P_{track} obtained for the considered ROI: the best deterministic approach (GaussFit&GNN) achieves a tracking ac-

Table 5.2: Results for real microscopy image sequences (P_{track} [%]). The mean value and standard deviation over all sequences for each of the eight approaches are also given.

	SEF& GNN	SEF& Kalman	SEF& MPF	SEF& IPF	GaussFit& GNN	GaussFit& Kalman	GaussFit& MPF	GaussFit& IPF
Seq. 1	75	81	85	87	71	81	82	83
Seq. 2	61	69	72	84	31	45	63	85
Seq. 3	45	32	77	43	41	60	60	80
Seq. 4	62	69	29	71	71	75	24	83
Seq. 5	94	94	85	94	81	94	86	94
Seq. 6	58	75	40	64	55	68	42	62
Seq. 7	65	81	48	82	53	77	31	78
Seq. 8	75	74	51	75	75	74	49	68
Seq. 9	73	83	49	74	68	81	53	77
Mean	68	73	60	75	61	73	54	79
Std. Dev.	14	17	21	15	17	14	21	9

curacy of 50%, whereas the best approach based on Kalman filters (GaussFit&Kalman) attains a tracking accuracy of 58%. The best approach based on a mixture of particle filters (GaussFit&MPF) achieves a tracking accuracy of 54%. In contrast, the best approach based on independent particle filters (GaussFit&IPF) obtains a tracking accuracy of 90%, thereby outperforming the other approaches.

For the entire image sequence "Seq. 4" (see Table 5.2, fourth row), the results are as follows: the best deterministic approach (GaussFit&GNN) achieves a tracking accuracy of 71%, while the best approach based on Kalman filters (GaussFit&Kalman) attains an accuracy of 75%. The best approach based on a mixture of particle filters (SEF&MPF) attains an accuracy of 29%, whereas the approach based on independent particle filters (GaussFit&IPF) yields an accuracy of 83%. The reason for the large performance difference between the approaches using MPF and the approaches using IPF is twofold. First, consider that the real images display a variable number of viruses. Since MPF cannot track a variable number of objects, its performance is diminished. In contrast, IPF is designed to handle an unknown and variable number of objects. Second, consider the fact that MPF employs a reclustering algorithm to allocate a set of particles to each of the components (objects) constituting the mixture model. Thus, some components may be allocated fewer particles, thereby attaining worse estimation results for these components. Instead, IPF uses a constant number of particles for each object, which leads to a uniform performance for all independent filters. Additionally, for MPF, in cases where virus particles lie in close proximity, the reclustering algorithm might yield non-compact clusters of particles, which leads to inaccurate position esti-

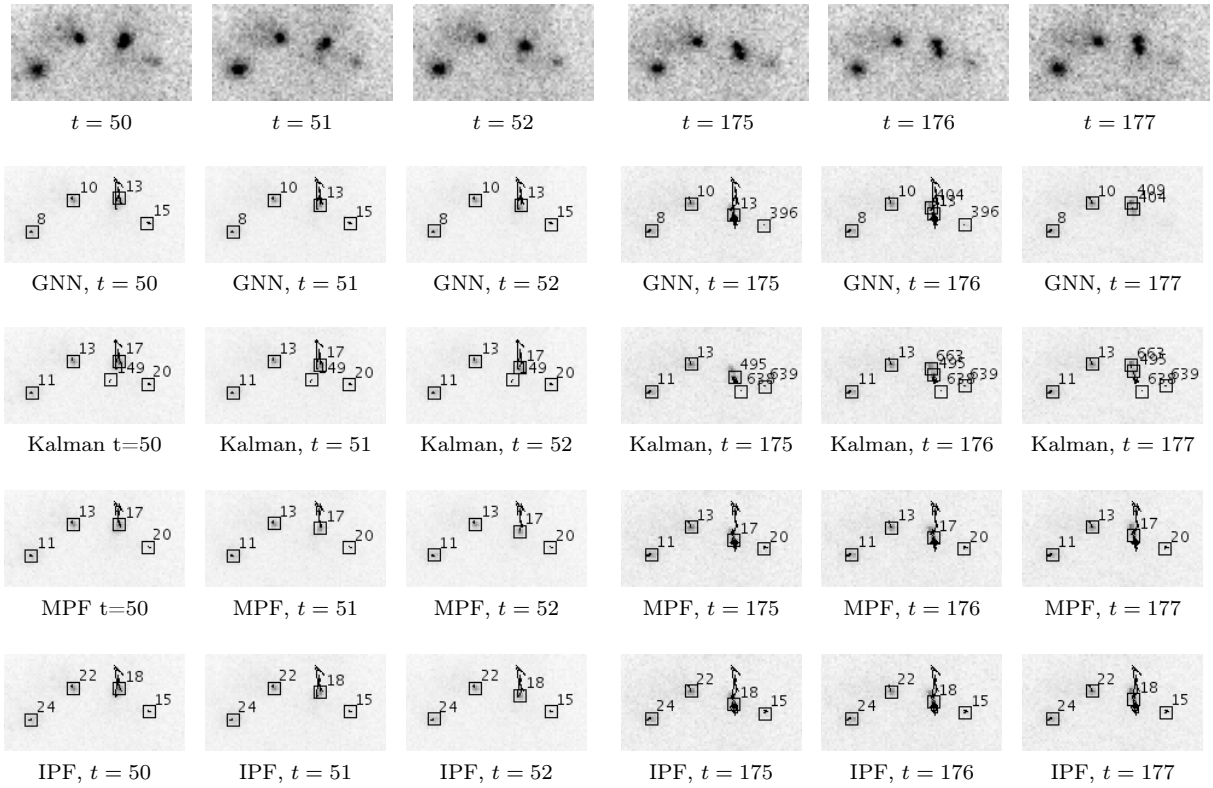


Figure 5.3: Original images (section, first row) and tracking results (second to fifth row) for the real image sequence "Seq. 1". For all shown results 2D Gaussian fitting was used for particle localization. (Image intensities have been inverted for visualization purposes).

mates. In contrast, IPF resolve such situations by means of our penalization scheme, thereby exhibiting an improved accuracy.

Another evaluated image sequence ("Seq. 1") consists of 250 images (256×256 pixels, 16-bit) and includes 23 well-defined virus particles. The experimental results for this sequence are listed in the first row of Table 5.2. Figure 5.3 shows some sample results for the approaches using 2D Gaussian fitting. It can be seen that both the MPF and IPF maintain the correct identity of the two rightmost particles, while the remaining approaches fail in this aspect. The reason for this result is that the deterministic approaches are easily misled by spurious objects. Similarly, the failures for the Kalman filter arise by incorrect object-measurement assignments due to spurious objects. In contrast, the particle filter is by design capable of handling various plausible hypotheses at one time. For this sequence, the best deterministic approach (SEF&GNN) attains a tracking accuracy of 75%, while the best approach based on Kalman filters (GaussFit&Kalman) obtains 81%. The best approach based on a mixture of particle filters (SEF&MPF) attains 85%, and the best approach based on independent particle filters (SEF&IPF) achieves an accuracy of 87%.

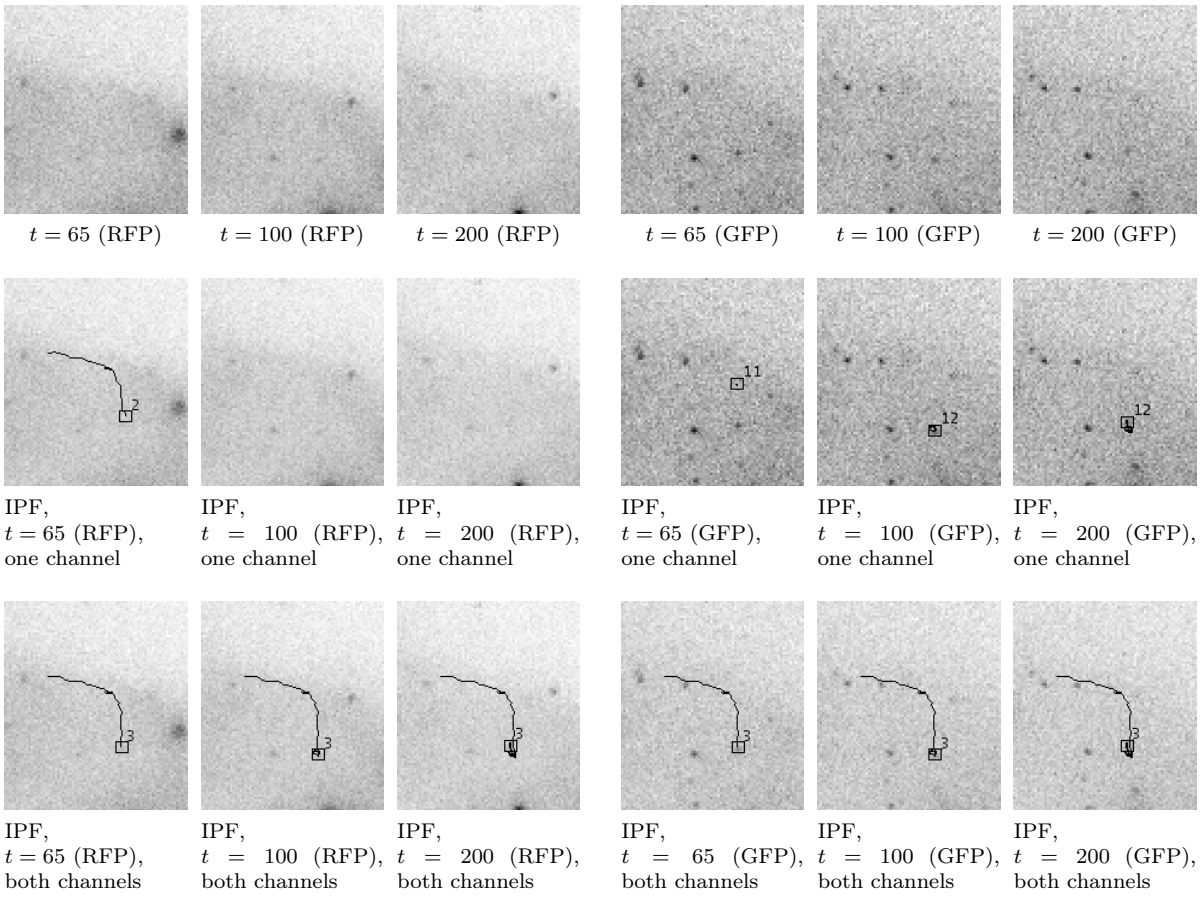


Figure 5.4: Original images (section) and tracking results for the real multi-channel image sequence "Seq. 5". The original images from the red fluorescent protein (RFP) (left) and green fluorescent protein (GFP) (right) channels are displayed in the first row. Tracking results for separate channels using IPF are shown in the second row; only the trajectories for the considered object are displayed. Tracking results when exploiting both channels using IPF are shown in the third row; only the trajectory of particle '3' is displayed. 2D Gaussian fitting was used for particle localization. (Image intensities have been inverted for visualization purposes).

The results obtained for all eight approaches applied to all nine image sequences are presented in Table 5.2. The mean values and standard deviations over all sequences are also given for each approach. It can be seen that the deterministic approaches perform quite adequately. The approach SEF&GNN, by employing simpler criteria for the rejection of noise-induced detections than GaussFit&GNN, yields fewer detection failures, which, in general, entail an improved tracking accuracy (mean tracking accuracy $\bar{P}_{track}$ of 68%). However, these simple criteria lead to the detection of spurious objects as well. On the other hand, the approach GaussFit&GNN employs stricter rejection criteria, thereby rejecting not only spurious objects but also real virus particles. The detection failures lead to broken trajectories, which result in a slightly lower tracking

accuracy ($\bar{P}_{track} = 61\%$). The approaches based on the Kalman filter tend to yield a relatively good tracking accuracy, which is higher compared to the corresponding deterministic approaches: both SEF&Kalman and GaussFit&Kalman attain $\bar{P}_{track} = 73\%$. These results agree well with our initial contention, namely that a spatial-temporal filtering step ought to enhance the overall tracking performance.

The approaches based on particle filters yield a fairly good tracking accuracy (see columns 4, 5, 8, and 9 in Table 5.2). It turns out that the approaches using IPF yield a significantly better result than the approaches using MPF. The reason for this difference, as discussed above, is twofold: first, the MPF cannot track a variable number of objects; second, MPF allocates a variable number of samples to each component (object), which results in poor estimation results for components with few particles. In contrast, IPF is able to cope with an unknown and variable number of objects. Additionally, IPF uses a constant number of samples for each object. Furthermore, our penalization scheme enables IPF to track successfully objects in close proximity. In fact, among all eight tracking approaches, the best mean tracking accuracy is obtained by GaussFit&IPF.

In summary, the results shown in Table 5.2 indicate that the mean accuracy $\bar{P}_{track}$ for the best deterministic approach (SEF&GNN) is 68%, while the mean accuracy of the best approach based on Kalman filters (SEF&Kalman) is 73%. In contrast, the mean accuracy of the best approach based on particle filters (GaussFit&IPF) is 79%. Note that, the performance for the last four image sequences is relatively low because of the high level of cellular autofluorescence and the smaller size of the particles.

We have also applied our approaches to both channels of real two-channel microscopy image sequences. In these sequences, the first channel corresponds to the red fluorescent protein (RFP) signal while the second channel corresponds to the green fluorescent protein (GFP) signal. Generally, virus particles are visible in both channels. In some cases, a particle may disappear in one of the channels, while remaining visible in the other channel. In these cases, it is important to exploit the information from both channels. For two-channel microscopy image sequences, the deterministic tracking approaches (SEF&GNN as well as GaussFit&GNN) work as follows: for particle localization, a set of position estimates is obtained by applying a union operation to the two sets of position estimates obtained by applying the localization approaches to the individual channels; the motion correspondence algorithm (cf. Section 2) subsequently operates on this consolidated set of measurements. Similarly, the Kalman filter operates on this set of position estimates as well. For the approaches using particle filters, we also use the consolidated set of position estimates and employ the multi-channel measurement model described in Section 3.8.1.5. In Figure 5.4, tracking results obtained with GaussFit&IPF for an ROI of 76×87 pixels for image sequence "Seq. 5" are shown.

Note the relatively high level of noise as well as the non-homogeneous background. Such a background generally poses a challenge to particle localization schemes based on intensity thresholds. However, our particle localization algorithms, in particular, the 2D Gaussian fitting approach, can cope well with such situations. The tracking results demonstrate that if only the information of the individual channels is employed, broken or incomplete trajectories are obtained for the virus particle in the middle of the ROI. In contrast, by employing the information from both channels, our approaches (in this case GaussFit&IPF) can retrieve the entire trajectory of this virus particle.

The computation times for the tracking approaches are as follows. For a synthetic image sequence encompassing 100 frames (64×64 pixels) with five objects, the computation time for the deterministic approach SEF&GNN is below 2 s, while GaussFit&GNN takes 11.06 s. The computation time for the probabilistic approach SEF&Kalman is also below 2 s, whereas GaussFit&Kalman requires 11.25 s. The particle filter approach GaussFit&MPF takes 1 min 33 s, while SEF&MPF requires 2 min 33 s. GaussFit&IPF requires about 30 s, while SEF&IPF takes about 1 min. For a real image sequence of 200 frames (512×512 pixels) with about 20 objects, SEF&IPF requires 14 min 43 s while GaussFit&IPF takes almost 3 h. All approaches were implemented in Java, except 2D Gaussian fitting, which was implemented in C++. The approaches have been implemented within our software named "ViroTracker" and were executed on a PC with an AMD Opteron (2.4 GHz) CPU running Linux.

5.1.2 Performance Evaluation of PDAE Approach Based on 2D and 3D Images

In the section above we have evaluated deterministic approaches and approaches based on the Kalman filter and the particle filter. There it turned out that the approach based on independent particle filters (IPF) yields the best result. However, the computational costs of the IPF are relatively high for applications in biological studies involving a large number of particles (e.g., 500 particles). In this section, we evaluate the proposed PDAE approach, which embraces the principles underlying the particle filter (e.g., multiple measurements, recognition-by-synthesis, use of parametric appearance models, direct use of the image data) in a more efficient manner. In this section, we present a quantitative evaluation of the PDAE approach and a comparison with approaches based on the Kalman filter and the particle filter. The evaluation is based on synthetic and real 2D and 3D image data.

Table 5.3: Summary of the evaluated probabilistic approaches.

	Kalman Filter	**IPF**	**PDAE**
Bottom-up detection	SEF	SEF	SEF
Motion correspondence	GNN	GNN	GNN
Number of measurements	1	N_s	N_m
Measurement source	SEF	$q_{\text{mix}}(\cdot)$	$\mathcal{N}(\cdot; \hat{\mathbf{p}}, \mathbf{S}_p)$ and $\mathcal{N}(\cdot; \mathbf{p}_{\text{BU}}, \mathbf{R}_p)$
Search strategy	Exhaustive	Random	Ellipsoidal
Measurement evaluation	–	$p(\mathbf{y}\|\mathbf{x})$	$p(\mathbf{z}\|\mathbf{x})$

5.1.2.1 Evaluated Probabilistic Tracking Approaches

We compare experimentally the performance of our approach based on probabilistic data association and an ellipsoidal sampling scheme (PDAE) with the performance of the standard Kalman filter (Section 3.5) and the performance of the approach based on independent particle filters (IPF) (Section 3.6). The Kalman filter as well as the IPF use a random walk model in 2D and 3D as described in Section 3.8.1. For the PDAE approach we also use a random walk model (see Section 3.8.1) or a random walk model as well as a directed motion model (see Section 3.8.3). For efficiency reasons, we used the spot-enhancing filter for bottom-up localization. For numerical stability, we use $N_s = 1000$ (2D case) or $N_s = 8000$ (3D case) samples for each independent particle filter. For the PDAE approach, we use $N_c = 4$ concentric contours and $N_j = 16$ positions along each elliptical contour. In 3D, we additionally evaluate concentric elliptical contours at $N_k = 8$ positions in z-direction. Together with the ellipsoidal measurements generated using the bottom-up measurement we obtain $N_m = 130$ measurements (samples) per filter (2D case) or $N_m = 1026$ measurements (3D case). Table 5.3 summarizes the evaluated probabilistic tracking approaches and their properties.

5.1.2.2 Evaluation on 2D and 3D Synthetic Images

We evaluate the performance of the approaches using synthetic image sequences. We consider two experimental scenarios. In the first scenario, we evaluate the localization accuracy of the approaches as a function of the image noise. In the second scenario,

we evaluate the tracking accuracy of the approaches as a function of the density of the objects. In the synthetic images we render each object using either a 2D Gaussian function (3.58) or a 3D Gaussian function (3.59) as the appearance model of the rendered objects.

To evaluate the localization accuracy, here we also calculated the root mean square error (RMSE) between the calculated positions $\mathbf{p}_{t,\text{calc}}$ and the true positions $\mathbf{p}_{t,\text{true}}$ of a true trajectory whose positions are indexed in time by $\mathcal{T} = \{\ldots, t-1, t, t+1, \ldots\}$ spanning $|\mathcal{T}| = N_{\text{steps,true}}$ time steps:

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{steps,true}}} \sum_{t \in \mathcal{T}} |\mathbf{p}_{t,\text{calc}} - \mathbf{p}_{t,\text{true}}|^2} \tag{5.3}$$

5.1.2.2.1 First Synthetic Scenario In this scenario we evaluate the localization accuracy of the approaches as a function of the image noise. The image noise is reflected by the signal-to-noise ratio (SNR), which is defined as the difference between the peak intensity $I_{\max}$ of an object and the intensity of the background I_b, divided by the standard deviation of the noise level σ_n ([36]). We assume a Poisson distribution for the noise model. We set the background intensity to $I_b = 10$ and vary the peak intensity $I_{\max}$ to explore different SNR levels. We explore the following SNR levels: 11.6, 8.8, 6.5, 4.6, 3.5, 2.8, 2, and 1.3. For each level we generate 30 image sequences. In 2D, each image sequence has 50 time steps. The image dimensions are 64×64 pixels. In 3D, each image sequence has 30 time steps and the image dimensions are 64×64×32 voxels. Each image sequence displays one object. The object is initially positioned at the center of the image. The motion of the object is governed by random walk, where the expected magnitude of the displacement over two time steps is set to 1.3 pixels. The true trajectory of the object is the same over all SNR levels. Thus for each SNR level only the (random) noise configuration varies over the corresponding 30 image sequences. For the appearance model, we use a value of $\sigma_{xy} = 2$ pixels and a value of $\sigma_z = 1$ pixel. The parameters of the tracking approaches are defined as follows. For the detection scheme based on the SEF, we adjust the factor c in (3.1) according to the SNR level and use a standard deviation of $\sigma_{\text{F},xy} = 1.5$ pixels in 2D and a standard deviation of $\sigma_{\text{F},z} = 1.5$ voxels in 3D. For all tracking approaches, the expected deviations for the position variables are fixed to $q_x = q_y = q_z = 4$ pixels^2. For the appearance variables, we use the following values: $q_{I_{\max}} = 1$ intensity level unit^2, $q_{\sigma_{xy}} = q_{\sigma_z} = 0.01$ pixels^2. The measurement errors take the following values: $r_x = r_y = r_z = 0.01$ pixels^2, $r_{I_{\max}} = 1$ intensity level unit^2, $q_{\sigma_{xy}} = q_{\sigma_z} = 0.36$ pixels^2. The threshold for the validation region is set to $\gamma_p^2 = 5.99$ in 2D and $\gamma_p^2 = 7.82$ in 3D. The expected degree of noise σ_n is set according to the SNR level.

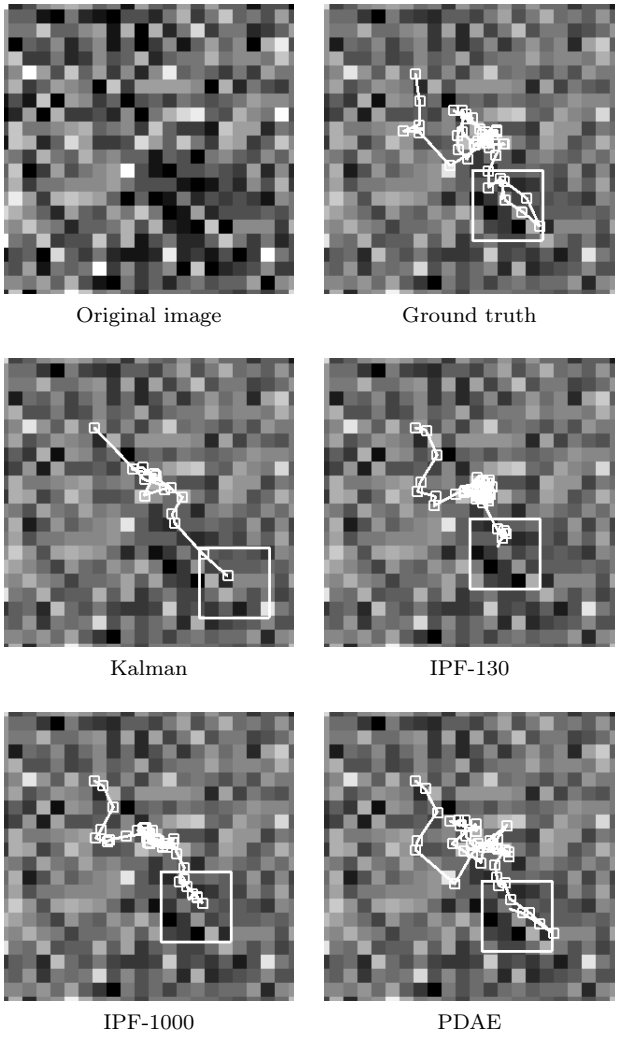
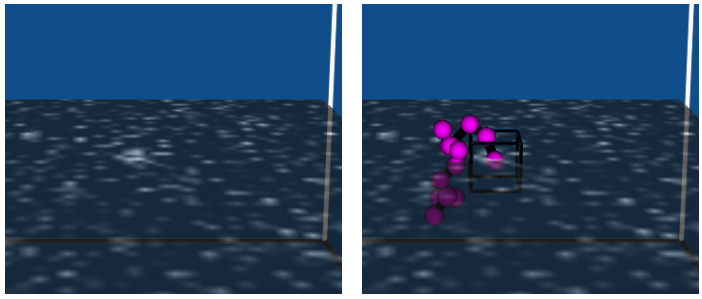


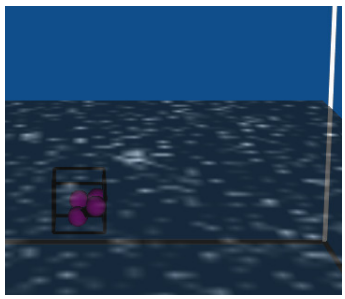
Figure 5.5: Original image, ground truth, and tracking results for the evaluated approaches on synthetic images. The SNR is 1.3 and the time step is $t = 39$. The small rectangles along the trajectories indicate the intermediate positions while the large rectangle indicates the current position. Image intensities have been inverted.

In this scenario we examine the performance of the PDAE approach in comparison to the IPF approach using the same number of samples N_s as the number of measurements N_m in the PDAE approach, viz. $N_s = 130$ in 2D and $N_s = 1026$ in 3D. We refer to these approaches as IPF-130 (2D case) and IPF-1026 (3D case), respectively; we refer to the approaches with higher number of samples as IPF-1000 (2D case) and IPF-8000 (3D case). Sample results for all approaches at SNR = 1.3 are shown in Figure 5.5 and Figure 5.6 for 2D and 3D images, respectively. It can be seen that the accuracy of the PDAE approach is superior to that of the Kalman filter at this challenging SNR level. The performance of the PDAE approach is quite similar to that of the particle filter using a large number of samples. Figures 5.7 and 5.8 show the results of all

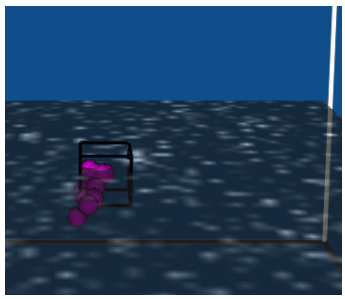


Original image

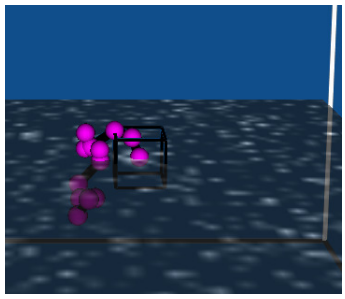
Ground truth



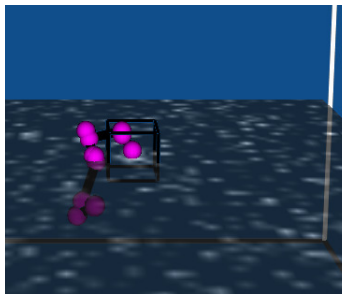
Kalman



IPF-1026



IPF-8000



PDAE

Figure 5.6: Original image, ground truth, and tracking results for the evaluated approaches on synthetic 3D images. The SNR is 1.3 and the time step is $t = 12$. A z-slice ($z = 13$) of the original volume image as well as the ground truth are shown. Trajectories are rendered as spheres (positions) and sticks (displacement vectors); the cube shows the current position.

evaluated approaches in terms of their RMSE as a function of the SNR level. The mean values of the RMSE computed over 30 image sequences corresponding to each SNR level are shown. For both 2D and 3D images, the PDAE approach outperforms the standard Kalman filter. This shows that the PDAE measurement process (cf. Section 3.7.1) improves the localization accuracy in comparison to using bottom-up particle localization schemes only. Also, for both 2D and 3D images, the PDAE approach outperforms the particle filter using the same number of samples ($N_s = 130$ in 2D and $N_s = 1026$ in 3D). This indicates that for the IPF a relatively low number of samples is not sufficient to obtain an accurate numerical approximation of the posterior probabilities. The PDAE approach is not subject to such numerical issues since the

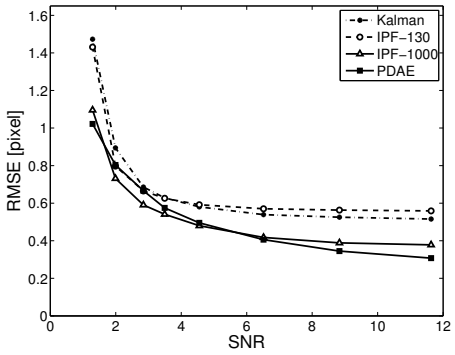


Figure 5.7: Tracking accuracy (RMSE) as a function of the SNR for 2D synthetic image sequences. The mean values computed over 30 image sequences at each SNR level are shown.

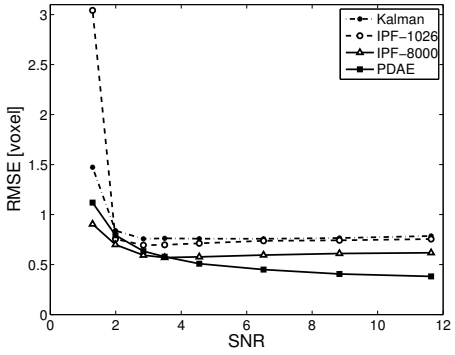


Figure 5.8: Tracking accuracy (RMSE) as a function of the SNR for 3D synthetic image sequences. The mean values computed over 30 image sequences at each SNR level are shown.

posterior probabilities are represented analytically via the underlying Kalman filter. Improved results are obtained for the IPF by using a larger number of samples (i.e., $N_s = 1000$ in 2D or $N_s = 8000$ in 3D). At the higher SNR levels, the PDAE approach outperforms the IPF with a larger number of samples. A reason for this is that the proposed ellipsoidal sampling scheme explores the position space systematically while the IPF explores this space randomly. The systematic approach yields a comparable performance relative to that of the random scheme with a *large* number of samples at the lower SNR levels. Since both approaches query the same image likelihood, the computation time of both approaches is linearly dependent on the number of samples (measurements). Thus the computational cost entailed by the PDAE approach is ca. 7.7 times lower than that of the IPF. This is particularly beneficial when tracking a large number of objects.

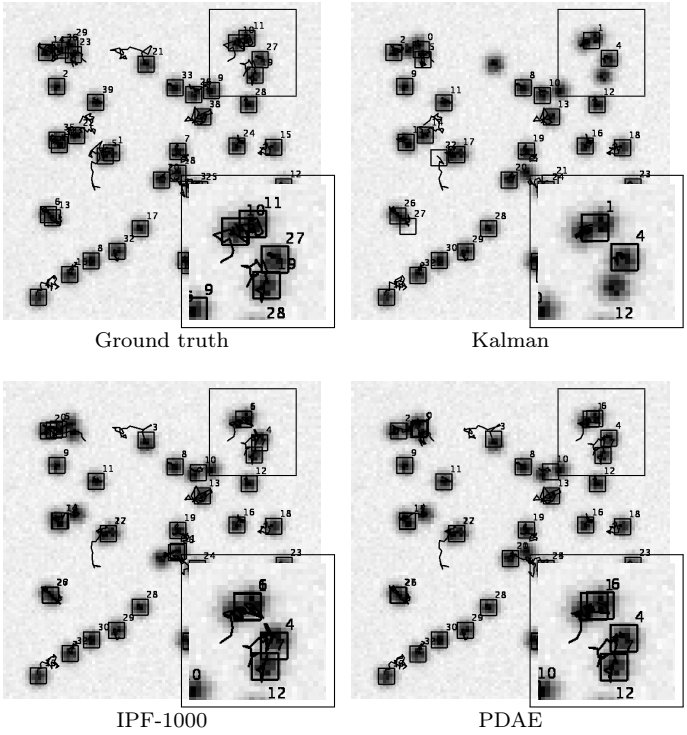


Figure 5.9: Ground truth and tracking results for all evaluated approaches on synthetic images. The probability of overlap of the objects is $p_{\text{overlap}} = 0.11$ and the time step is $t = 27$. The numbers label each trajectory. Insets show the tracking results of the region enclosed with a rectangle at a higher resolution.

5.1.2.2.2 Second Synthetic Scenario We evaluate here the performance of the PDAE approach as a function of the object density. The object density d_{obj} is given by the ratio between the number of objects N_{obj} and the size (area) of the image S_{image}. Each object occupies a certain area S_{obj}. The definition of d_{obj}, however, does not take into account S_{obj}. To take into account the area occupied by each object, we calculate the probability of overlap p_{overlap} [172], which is defined as:

$$p_{\text{overlap}} = \frac{N_{\text{obj}} S_{\text{obj}}}{S_{\text{image}}}. \tag{5.4}$$

Thus, the higher the number of objects, the higher the probability of overlap among objects, and therefore the more tracking errors occur. In this scenario we generate 2D image sequences consisting of 30 time steps. The images have dimensions 100×100 pixels (16-bit) and thus $S_{\text{image}} = 10000$ pixels^2. We use the following number of objects: $N_{obj} = 10$, 20, 30, 40, 50, 75, and 100. We use the same appearance parameters as in the first scenario for rendering all objects. Assuming that each Gaussian object with $\sigma_{xy} = 2$ pixels can be described by a 2D disk with radius $r = 1.5\sigma_{xy} = 3$ pixels, the area of each object is given by $S_{\text{obj}} = r^2\pi = 28$ pixels^2 and thus we explore the following

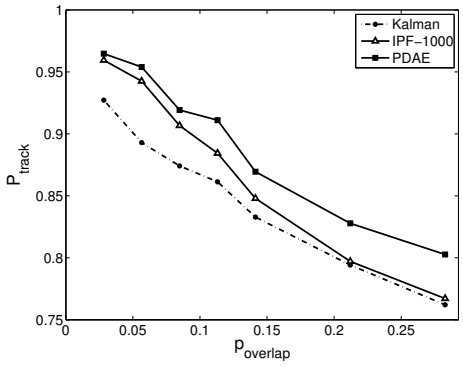


Figure 5.10: Tracking accuracy (P_{track}) as a function of the probability of overlap of the objects p_{overlap}. The mean values computed over 30 image sequences for each probability of overlap are shown.

values for the probability of overlap p_{overlap}: 0.03, 0.06, 0.08, 0.11, 0.14, 0.21, 0.28. The SNR is 11.6. The initial image position of each object is random and the motion is governed by random walk. The expected deviation for the displacement takes values of up to 1.3 pixels. For each number of objects N_{obj}, we generate 30 image sequences. The factor c for the detection scheme based on the SEF is adjusted according to the number of objects, since the image statistics (mean intensity and standard deviation) vary with the number of objects. For all tracking approaches, the expected deviations for the position variables are fixed to $q_x = q_y = q_z = 0.56$ pixels^2. Trajectories are initialized only at the first time step. All other parameters are the same as in the first synthetic scenario.

Results for all approaches at a probability of overlap $p_{\text{overlap}} = 0.11$ (number of objects $N_{\text{obj}} = 40$) are presented in Figure 5.9. The results show that the PDAE approach preserves the identity of neighboring objects relatively well. The performance of the approaches in terms of P_{track} is shown in Figure 5.10. The performance follows approximately the theoretical detection performance $1 - p_{\text{overlap}}$ [172]. It can be seen that the PDAE approach consistently outperforms the other approaches. This shows the viability of the image support scheme that computes the support at each image position for tracking multiple objects. We also measured the computation time of each approach. All approaches were implemented in Java within our software ViroTracker [70] which was executed on an Intel Xeon X5650 (6 cores at 2.6 GHz) machine running Linux. Figure 5.11 shows the computation time as a function of the number of objects (mean computation time over 30 2D image sequences with 30 time steps each.) The approach with the lowest computation time is the Kalman filter. For the IPF approach and the PDAE approach, the computational cost is linearly dependent on the number

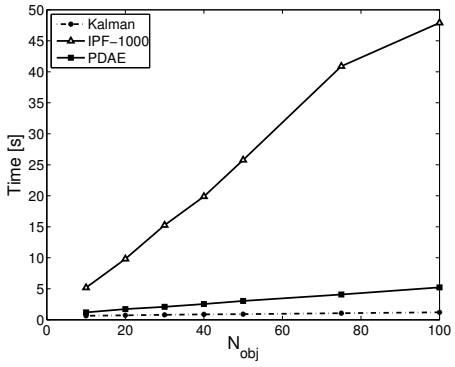


Figure 5.11: Computation time (s) as a function of the number of objects. The mean values computed over 30 2D image sequences consisting of 30 time steps for each number of objects are shown.

of samples (measurements) per object. In comparison to the Kalman filter, the PDAE approach is computationally somewhat more expensive (approximately a factor of 3). In comparison to the IPF with 1000 samples, the PDAE approach using 130 measurements is about 7.7 times faster, as expected. For example, for $N_{\text{obj}} = 40$, the PDAE approach entails a mean computation time of 2.5 s while the IPF requires about 20 s. Thus, in comparison to the Kalman filter, the PDAE approach yields a superior performance at the expense of a somewhat higher computational cost. In comparison to the IPF, the PDAE approach delivers also a better performance in terms of P_{track} at a significantly lower computational cost.

5.1.2.3 Evaluation on 2D and 3D Real Microscopy Images

5.1.2.3.1 2D Images We have applied our PDAE tracking approach to the nine real microscopy 2D image sequences described in Section 5.1.1.3. For the spot detection scheme we use a standard deviation of $\sigma_{\text{F},xy} = 1.5$ pixels. For the adaptive threshold, we use a factor of $c = 2.7$ for the first four image sequences; for the other image sequences we use $c = 3.7$. For the PDAE approach we use a random walk model as well as a directed motion model. For the random walk motion model, we use a value of $q_x = q_y = q_z = 4$ pixels^2. For the directed motion model, we use a value of $q_{\text{CV}} = 4$ pixels^2/frame^3. The noise parameter σ_n is adjusted based on the measured peak intensity for the particles. We performed a comparison with the Kalman filter and the IPF. For the Kalman filter, the distance metric $d(\cdot, \cdot)$ used in the motion correspondence step is in this case defined by the Mahalanobis distance using the predicted measurement covariance $\mathbf{S}$ in (3.12). We set the maximum expected distance to $d_{max} = 7.82$. The disappearance duration is set to 20 frames for both the approach based on the standard Kalman filter as well as

Table 5.4: Results for 2D real image sequences in terms of the tracking accuracy P_{track} [%]. The mean values and standard deviations are also shown.

Sequence	Kalman	IPF-1000	PDAE
1	96	96	98
2	81	90	92
3	90	92	99
4	82	82	89
5	94	94	94
6	68	68	75
7	86	87	88
8	77	78	79
9	81	83	84
Mean	84	85	89
Std. Dev.	9	9	8

for the PDAE approach. All other parameters including those for the IPF are set as in the first synthetic scenario (cf. Section 5.1.2.2.1).

Sample results for all approaches are shown in Figure 5.12. In this example, the particle alternates between random motion and directed motion. The Kalman filter and the IPF approach yield broken trajectories. In comparison, the PDAE approach obtains a trajectory without gaps that is very similar to the ground truth trajectory. For all approaches, Table 5.4 shows the quantitative results for all nine image sequences in terms of P_{track}. The Kalman filter yields P_{track}= 84% (standard deviation of 9%) while the IPF yields a mean accuracy of P_{track}= 85% (standard deviation of 9%). The PDAE approach yields the best results with a mean accuracy of P_{track}= 89% (standard deviation of 8%).

5.1.2.3.2 3D Images We have also applied the approaches to a real microscopy 3D image sequence. The image sequence displays budding HIV-1 particles [90]. The particles are labeled with eGFP. Images were acquired with a Nikon TE2000-E spinning disk confocal microscope using an Andor Technology EM-CCD camera. The image sequence consists of 395 time steps. For each time step, a stack of 7 images (slices) with dimensions 255 $\times$ 512 pixels was acquired. The ground truth for the 3D positions of the particles was obtained manually with the Manual_Tracking plug-in of ImageJ [1]. In total, 15 particles were manually tracked. For the spot detection scheme we use

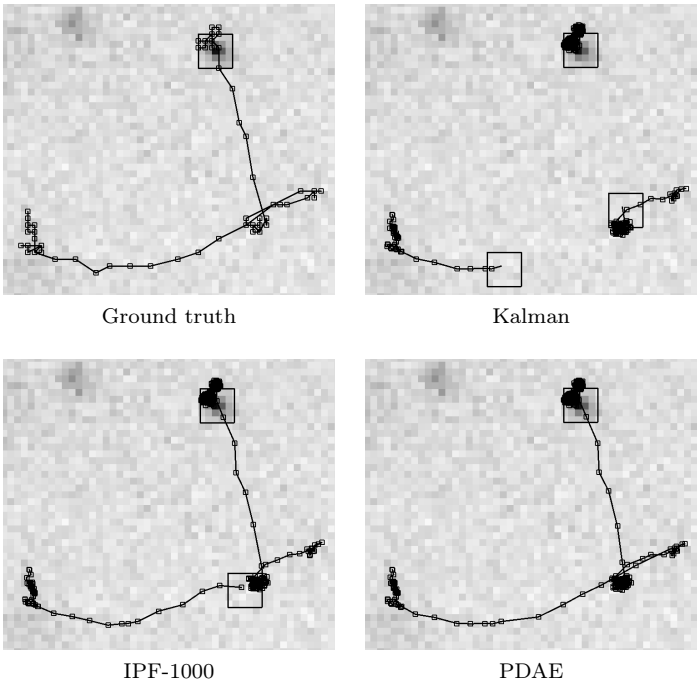


Figure 5.12: Ground truth and tracking results for the evaluated approaches on real 2D microscopy images (time step $t = 400$). The small rectangles along the trajectories indicate the intermediate positions while the large rectangle indicates the final position. Image intensities have been inverted.

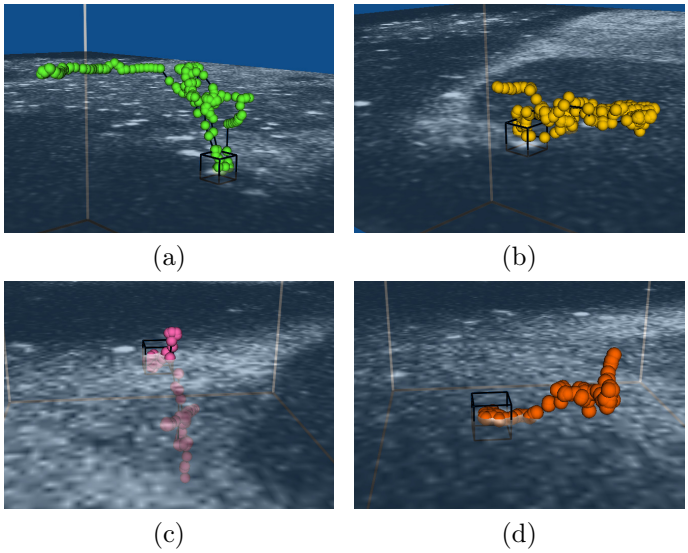


Figure 5.13: Tracking results of the PDAE approach on a real 3D microscopy image sequence (time step $t = 368$). Individual trajectories are shown. Z-slices of the original volume image are displayed. Small spheres along the trajectories represent the intermediate positions while cubes show the current position of the particles.

$\sigma_{\mathrm{LoG},xy} = 1.5$ voxels, $\sigma_{\mathrm{LoG},z} = 1$ voxel, and $c = 5$. All other parameters remain the same

Figure 5.14: Tracking results of the PDAE approach on a real 3D microscopy image sequence (time step $t = 368$). A z-slice ($z = 6$) of the original volume image is shown. Small spheres along the trajectories represent the intermediate positions while cubes show the current position of the particles. For visualization purposes, only the 50 previous intermediate positions are shown for each trajectory. Colors are used to distinguish the different trajectories.

as for the 2D image sequences.

The performance of all approaches is relatively good. In terms of the tracking accuracy P_{track}, the Kalman filter yields P_{track}=88%, the IPF yields P_{track}=90% while the PDAE approach yields P_{track}=91%. Analogous to the 2D images, the proposed approach performs better than the other approaches. In Figure 5.13 sample results for individual trajectories obtained with the PDAE approach are presented. In Figure 5.14 all trajectories computed by the PDAE approach are shown. It can be seen that the PDAE approach deals successfully with a large number of particles.

5.2 Behavior Identification: Experimental Results

We have evaluated the performance of the layered probabilistic approach for behavior identification described in Chapter 4 using synthetic image sequences as well as real microscopy image sequences [69].

5.2.1 Experimental Procedures

To track virus particles in two-channel image sequences, we use our probabilistic tracking approaches described in Chapter 3. Note that the image information from both channels is exploited simultaneously by using the measured position estimates computed on both channels. We consider three synthetic scenarios and two real scenarios. For all synthetic image sequences and for all real image sequences of the first scenario, we used

independent particle filters for tracking (see Section 3.6). For the real images of the second scenario, Kalman filters were used (cf. Section 3.5). Besides the positions of single particles, the tracking approach determines the intensity over time $y_{1:T}$. To reduce the influence of noise on the measured intensity, the intensity statistic is computed as a weighted mean of the intensities within a spatial neighborhood at the position of the particle, with weights according to a 2D Gaussian function centered at the position of the particle. Note that this approach for computing the intensity statistics is applicable to different localization schemes (spot detection schemes). Since the dimension of our state space is low, $N_s = 1000$ samples for the hybrid particle filter ensure a sufficient support of each posterior model probability. In our experiments, the transition matrix $\mathbf{\Pi}$ for the temporal intensity models takes the following form in accordance with the considerations in Section 4.5.1:

$$\mathbf{\Pi} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0.0 \\ 0.1 & 0.0 & 0.9 \end{pmatrix}. \tag{5.5}$$

Similarly, the transition matrix $\mathbf{\Phi}$ for the behaviors follows the considerations in Section 4.5.2 and is defined as follows:

$$\mathbf{\Phi} = \begin{pmatrix} 0.5 & 0.2 & 0.2 & 0.1 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.5 \end{pmatrix}. \tag{5.6}$$

Different values for the variances Q_α according to the different intensity statistics of each channel are specified. The initial prior probabilities $P(\alpha_0)$ for the three different intensity models are set to $P(\alpha_0) = [\ 0.7\ \ 0.15\ \ 0.15\ ]^T$. Since most particles do not fuse, the prior probabilities $P(\beta_0)$ for the four behaviors are set to $P(\beta_0) = [\ 0.5\ \ 0.2\ \ 0.2\ \ 0.1\ ]^T$.

For an experimental comparison with a previous approach, we implemented a recent derivative-based approach [61] that exploits the first derivative of the intensity over time. To compute the first derivative with respect to time, finite differences over a local temporal neighborhood are used. Changes in the intensity are detected based on a threshold T_{deriv} for the absolute value of the calculated derivative values. The sign of the derivative values indicates whether the change is positive or negative. Time steps associated with a positive or negative change are assumed to be governed by a positive intensity change (PIC) or a negative intensity change (NIC) model, respectively. Time

steps that are not associated with a change in intensity are assumed to follow a constant intensity (CI) model. The approach determines the underlying behavior of the particle by applying the composite function $(h^{-1} \circ g)(\boldsymbol{\alpha})$ (cf. Section 4.5.2) that maps the vector $\boldsymbol{\alpha}$ of temporal intensity models computed on both channels to a certain behavior β.

For each time step t, the evaluated approaches compute discrete values (i.e., labels) for the intensity models $\boldsymbol{\alpha}_t = (\hat{\alpha}_t^0, \hat{\alpha}_t^1)$ as well as for the behavior β_t. For each of these three labels, we compute the labeling accuracy P_{label}, which is given as the percentage of correctly labeled time steps relative to all time steps of a single trajectory. The labeling accuracy quantifies the performance over individual time steps. To obtain a measure for individual trajectories, we categorize each trajectory associated with a fusion behavior as a *fusion trajectory*; otherwise the trajectory is defined as a *non-fusion trajectory*. Based on these two categories, we compute the accuracy of behavior identification P_{ident}, the identification error E_{ident}, as well as the precision P_{pre}. The first performance measure P_{ident} reflects the percentage of correctly identified fusion trajectories (True Positives) relative to the number of true fusion trajectories (True Positive Rate or recall/sensitivity). The second measure E_{ident} is the percentage of false fusion trajectories (False Positives) relative to the number of true non-fusion trajectories (False Positive Rate). The precision P_{pre} is defined as the ratio between the number of correctly identified fusion trajectories (True Positives) and the number of identified fusion trajectories (True Positives and False Positives). The ground truth consists of a set of trajectories with labels for the intensity models as well as for the behaviors. Each trajectory is also designated as a fusion or a non-fusion trajectory. For the synthetic data, the known intensity statistics of the objects are used to generate the ground truth. For the real data, the ground truth is obtained by manual annotation.

5.2.2 Evaluation on Synthetic Images

We evaluate the robustness of our approach using synthetic two-channel image sequences. We compare the performance of our scheme with that of the derivative-based approach [61]. We investigate the following three scenarios. In the first scenario, the aim is to evaluate the performance of the layered approach for the task of detecting fusion at different levels of image noise. This scenario considers single objects. In the second scenario, we model multiple objects and the goal is to examine the capability of the layered approach to retrieve the number of objects undergoing fusion from a set of objects with heterogeneous behaviors. We also investigate different levels of image noise. In the third scenario, multiple objects are considered and we examine the performance of the layered approach as a function of the performance of the tracking scheme. In all scenarios we render individual objects using a realistic appearance model defined by a

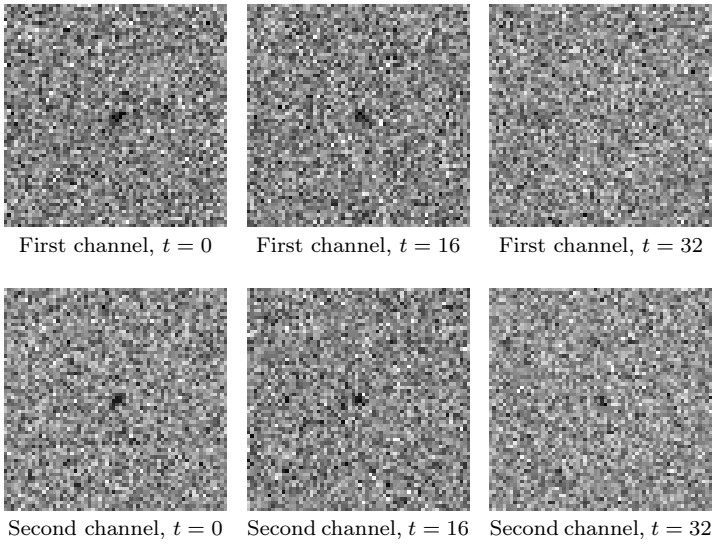
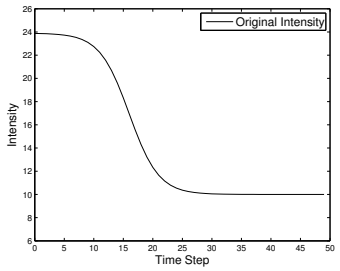


Figure 5.15: First synthetic scenario: Sample time steps of a synthetic image sequence. The SNR is 2.8. The intensity over time of the particle in the first channel is governed by a sigmoid model. The intensity over time in the second channel is given by an exponential model. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

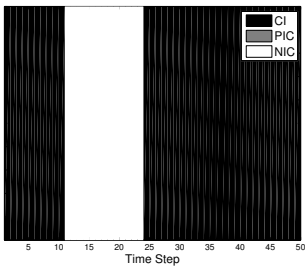
2D Gaussian function (see (3.58)). We define the signal-to-noise ratio (SNR) as the difference between the peak intensity I_{max} of an object and the intensity of the background I_b, divided by the standard deviation of the noise level σ_n ([36]). The noise model for the intensity is assumed to follow a Poisson distribution. To model different levels of SNR, we set the background intensity to $I_b = 10$ and vary the peak intensity I_{max}. In total, we explore seven SNR levels: 11.6, 8.8, 6.5, 4.6, 3.5, 2.8, and 2. In the first and second scenarios, the position (x, y) over time of individual particles is held constant. In the third scenario, the position over time of individual particles changes and is governed by random walk. The peak intensity I_{max} in each channel is varied over time according to the simulated behaviors of the different scenarios.

5.2.2.1 First Synthetic Scenario

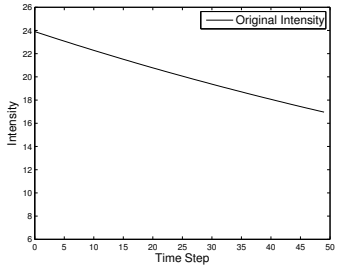
In the first synthetic scenario, we examine the performance of the layered approach for identifying the behavior of a single particle undergoing fusion at different SNR levels. For each of the seven SNR levels, we generate 30 two-channel image sequences. Each two-channel image sequence consists of 50 time steps, and each time step includes two images (one for each channel), where each image (16-bit) has dimensions 64×64 pixels. We consider a stationary object positioned at the center of the image. The reason for holding the position constant is to minimize tracking errors, which would otherwise bias the true performance of the layered approach for fusion detection. The peak intensity



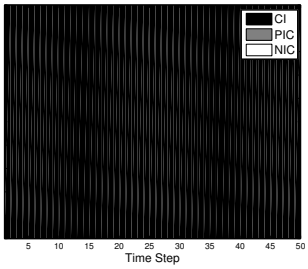
(a) Original intensity (first channel)



(b) True labels for the intensity models (first channel)



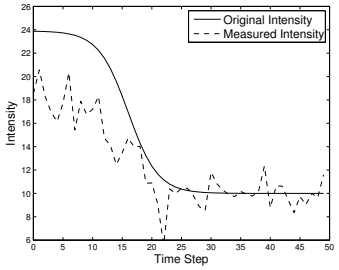
(c) Original intensity (second channel)



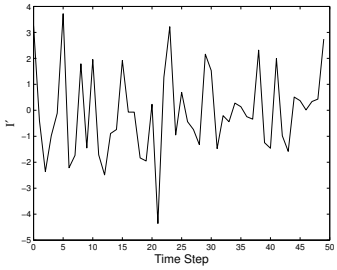
(d) True labels for the intensity models (second channel)

Figure 5.16: First synthetic scenario: Ground truth for the intensity as well as for the intensity models.

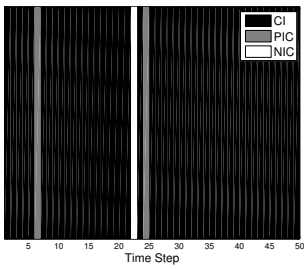
I_{max} of the object in the first channel is modeled using a sigmoid function (mirrored about the y-axis) with a steep transition, which simulates the rapid decrease of the intensity that characterizes the fusion behavior of real virus particles. The peak intensity I_{max} in the second channel follows an exponential function. To generate the ground truth labels for the temporal intensity models of each channel, we apply a threshold to the derivative of the underlying intensity model (i.e., the sigmoid or the exponential model). The ground truth labels for the temporal intensity models $\boldsymbol{\alpha}$ are used for determining the ground truth labels for the behavior β using the composite function $(h^{-1} \circ g)(\boldsymbol{\alpha})$ (cf. Section 4.5.2). For tracking, we use Gaussian fitting and independent particle filters on each two-channel image sequence. We apply the derivative-based approach as well as the layered approach to the synthetic images. For the derivative-based approach, the threshold values T_{deriv} are adjusted according to the SNR level. For the layered approach, the values for the variances Q_α that regulate the temporal intensity models as well as the noise parameter σ_n are also adjusted based on the SNR level. For both approaches, the parameter values for the first channel are determined empirically so that the approaches detect the negative change entailed by the sigmoid function without inducing too many incorrect labels for the intensity models. The parameter values for the second channel are chosen by taking into consideration that no significant changes in the intensity are



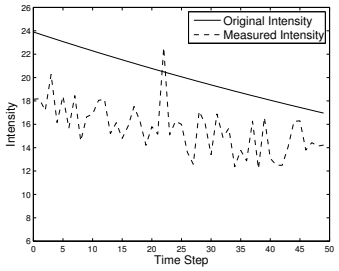
(a) Original and measured intensity (first channel)



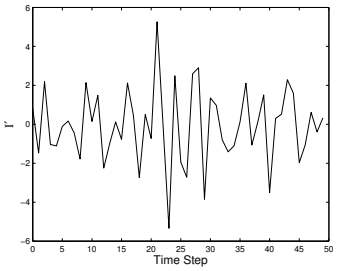
(b) Computed derivative of the measured intensity (first channel)



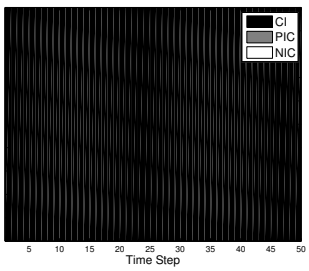
(c) Computed labels for the intensity models (first channel)



(d) Original and measured intensity (second channel)



(e) Computed derivative of the measured intensity (second channel)



(f) Computed labels for the intensity models (second channel)

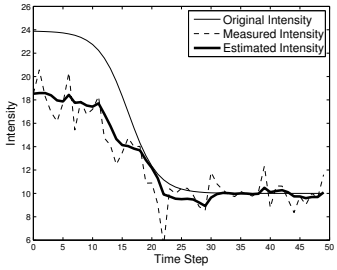
Figure 5.17: First synthetic scenario: Results obtained using a derivative-based approach. The SNR is 2.8. The labeling accuracy is $P_{label} = 72\%$ for the first channel and $P_{label} = 100\%$ for the second channel.

expected. For each SNR level, the parameter values are determined based on one image sequence; these values are then used for all 30 image sequences of the corresponding SNR level. The same parameter values are also used for the second and third synthetic scenarios (see Sections 5.2.2.2 and 5.2.2.3 below).

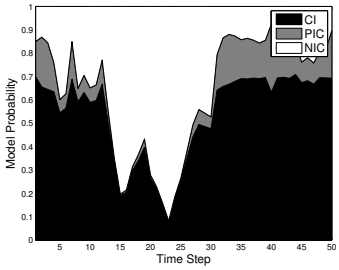
Figure 5.15 displays sample images from the two channels at a SNR of 2.8. Note the fast drop in the intensity for the object in the first channel. The original intensities are shown in Figure 5.16 together with the true labels for the intensity models of both channels. Although the intensity of the particle drops to the background level in the first channel, the tracking approach determines correctly the position by exploiting the image data from the second channel. The results for the derivative-based scheme, including the calculated derivative values as well as the resulting labels for the intensity models, are shown in Figure 5.17. For the first channel, one can see that the decrease in the intensity introduced by the sigmoid model is detected, although the period corresponding to this decrease is rather short. Because of the noise, positive changes in the intensity are also detected. The labeling accuracy (see Section 5.2.1) is $P_{label} = 72\%$ for the first channel while for the second channel the approach achieves a labeling accuracy of $P_{label} = 100\%$. Note that in Figure 5.17a the original intensity (solid line) represents the peak intensity of a particle while the measured intensity (dashed line) represents a spatial average

in the neighborhood of a particle (cf. Section 5.2.1), therefore the values are typically lower. The results obtained by our layered approach for both channels are displayed in Figure 5.18. The intensity for either channel is estimated well. The appropriate intensity models are activated as the intensity fluctuates and the computed sequence of intensity models agrees well with the sequence of true intensity models. Here the labeling accuracy is $P_{label} = 86\%$ for the first channel and $P_{label} = 100\%$ for the second channel. In comparison to the derivative-based approach, the layered approach yields a less fragmented result for the sequence of intensity models. Also the layered approach provides better results in the presence of noise. In Figure 5.19 we show the ground truth labels for the behavior. The labels computed by the derivative-based approach as well as by the layered approach are also shown. Both approaches detect the FUSION behavior, however, the layered approach reconstructs the original labels with a higher fidelity. A certain time lag between the original labels and the labels computed by the layered approach can be observed. This delay is due to the adaptation time of the hybrid particle filter and depends on the transition probabilities in $\mathbf{\Pi}$, the variances Q_α that regulate the temporal intensity models, as well as the noise parameter σ_n. While the adaptation time could be reduced, a compromise between the adaptation time and the steady behavior of the hybrid particle filter needs to be found. The quantitative values for the labeling accuracy confirm the superior performance of the layered approach, since the derivative-based approach achieves a labeling accuracy of $P_{label} = 76\%$ while the layered approach attains a labeling accuracy of $P_{label} = 90\%$.

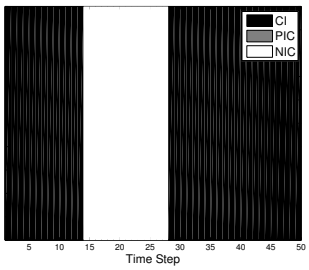
The labeling accuracy P_{label} of the evaluated approaches as a function of the SNR is presented in Figure 5.20. The diagrams display the mean and standard deviation of the labeling accuracy for the intensity models of the individual channels as well as for the behavior. For each SNR level, the mean and standard deviation are computed over 30 image sequences. For the first channel (see Fig. 5.20a), the layered approach yields better results than those delivered by the derivative-based approach. For the second channel (see Fig. 5.20b), the performance for both approaches is equally high. For the behavior (see Figure 5.20c), the labeling performance attained by the derivative-based approach is fair. Here, the performance for the behavior is bound by the labeling performance on the intensity models. In comparison, the top-level Markov chain of our layered approach copes well with potential errors in the computed intensity models. In general, the degradation of the performance of both approaches relative to the SNR is smooth and monotonic. This indicates that the SNR influences directly the performance of the approaches. In summary, the results show that our layered approach performs well for the task of behavior identification, achieving a fairly good performance at low SNR levels.



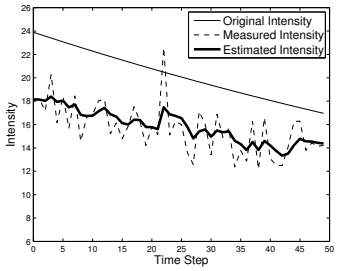
(a) Original, measured, and estimated intensity (first channel)



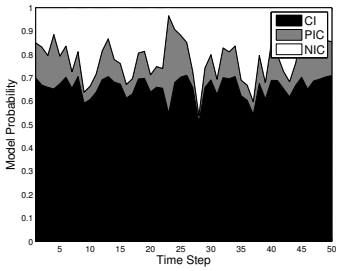
(b) Computed posterior model probabilities for the intensity models (first channel)



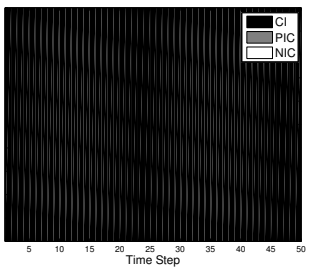
(c) Computed labels for the intensity models (first channel)



(d) Original, measured, and estimated intensity (second channel)



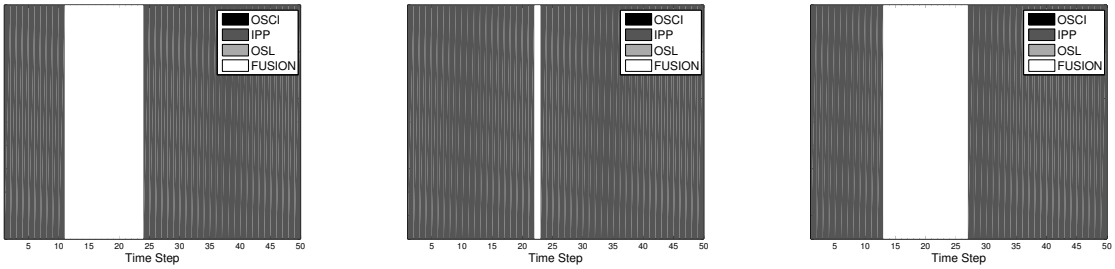
(e) Computed posterior model probabilities for the intensity models (second channel)



(f) Computed labels for the intensity models (second channel)

Figure 5.18: First synthetic scenario: Results obtained using a hybrid particle filter. The SNR is 2.8. The labeling accuracy is $P_{label} = 86\%$ for the first channel and $P_{label} = 100\%$ for the second channel.

Within this scenario we also evaluate the performance of the layered approach as a function of the number of samples N_s. We use the same 30 two-channel image sequences as above for SNR = 11.6 and evaluate the result for N_s = 30, 50, 100, 200, 500, and 1000. Figure 5.21a shows the mean and standard deviation of the labeling accuracy P_{label} for the first channel as a function of N_s. It can be seen that the performance of the approach is fairly good even for a relatively low number of samples (e.g., $N_s = 50$). The reason for this is that the dimension of the hybrid state space is relatively low. However, as indicated by the error bars, the standard deviation is relatively large for a low number of samples. Thus a larger number of samples (e.g., $N_s \geq 500$) is required to obtain robust results (in all our experiments we used $N_s = 1000$). A similar trend is observed for the labeling accuracy P_{label} for the second channel as well as for the behavior (see Figs. 5.21b, 5.21c). Figure 5.22 shows how the error $(1 - P_{label})$ decreases with an increasing number of samples N_s. Here N_s takes values between 10 and 200 with an increment of 10.

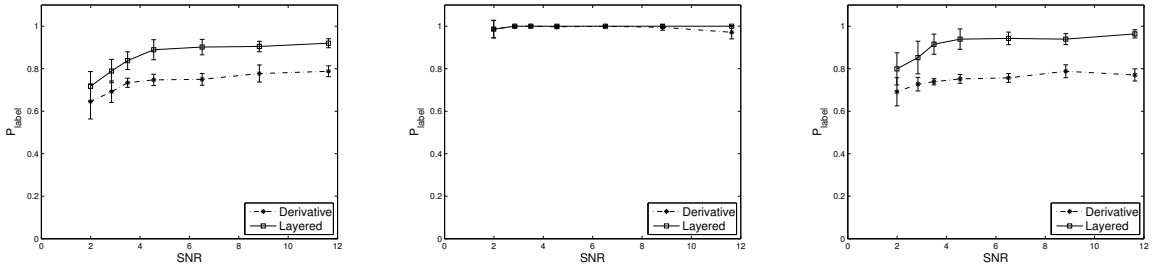


(a) True labels for the behavior

(b) Computed labels for the behavior using a derivative-based approach

(c) Computed labels for the behavior using the layered approach

Figure 5.19: First synthetic scenario: Ground truth for the behavior and results obtained using a derivative-based approach as well as the layered probabilistic approach. The SNR is 2.8. The labeling accuracy for the derivative-based approach is $P_{label} = 76\%$ while for the layered approach the labeling accuracy is $P_{label} = 90\%$.



(a) P_{label} for the intensity models (first channel)

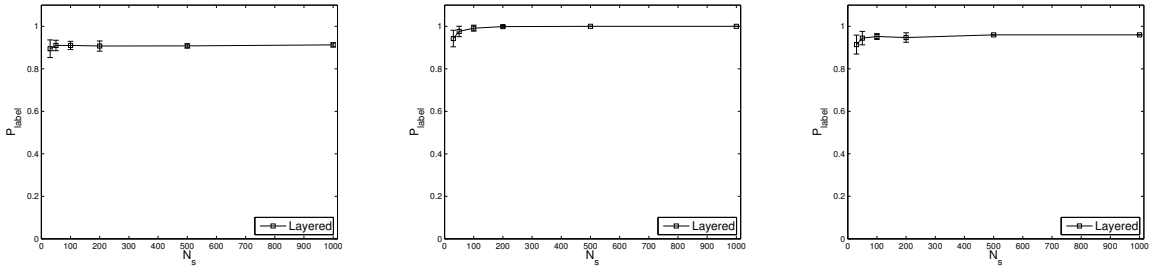
(b) P_{label} for the intensity models (second channel)

(c) P_{label} for the behavior

Figure 5.20: First synthetic scenario: Labeling accuracy P_{label} as a function of the SNR for the intensity models (first and second channel) as well as for the behavior. The mean values (and standard deviations) of a derivative-based approach ('Derivative') as well as the mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.

5.2.2.2 Second Synthetic Scenario

In the second scenario, we consider multiple objects and the goal is to study the performance of the layered approach for determining fusion given a set of particles with different behaviors. In this scenario, we generate two-channel image sequences consisting of 50 time steps. The images (256×256 pixels, 16-bit) display 30 objects rendered with a 2D Gaussian appearance model (see (2.35)). The image positions of the objects are randomly chosen and their positions remain constant over time. For each SNR level we generate 30 two-channel image sequences. The peak intensity I_{max} of each object in the first channel is modeled by either a constant, linear, exponential, or sigmoid function. In real microscopy images, photobleaching induces a slight decrease in the intensity. To simulate this phenomenon in our experiment, the linear and exponential

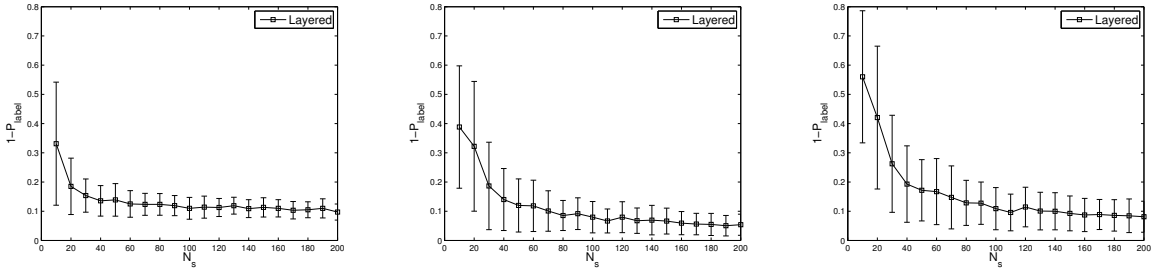


(a) P_{label} for the intensity models (first channel)

(b) P_{label} for the intensity models (second channel)

(c) P_{label} for the behavior

Figure 5.21: First synthetic scenario: Labeling accuracy P_{label} as a function of the number of samples N_s for the intensity models (first and second channel) as well as for the behavior. The mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.



(a) Error for the intensity models (first channel)

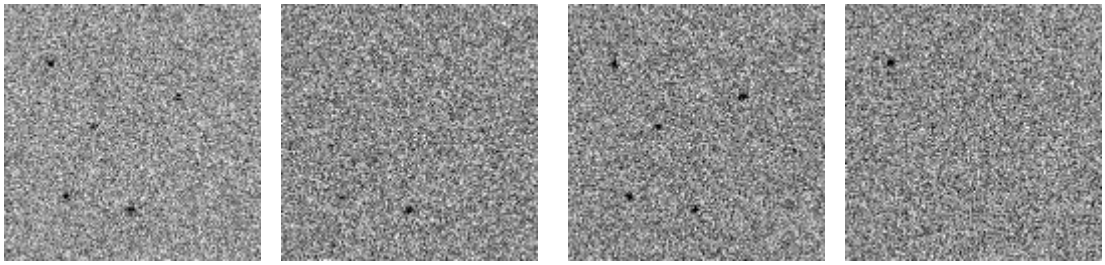
(b) Error for the intensity models (second channel)

(c) Error for the behavior

Figure 5.22: First synthetic scenario: Error $(1 - P_{label})$ as a function of the number of samples N_s for the intensity models (first and second channel) as well as for the behavior. The mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.

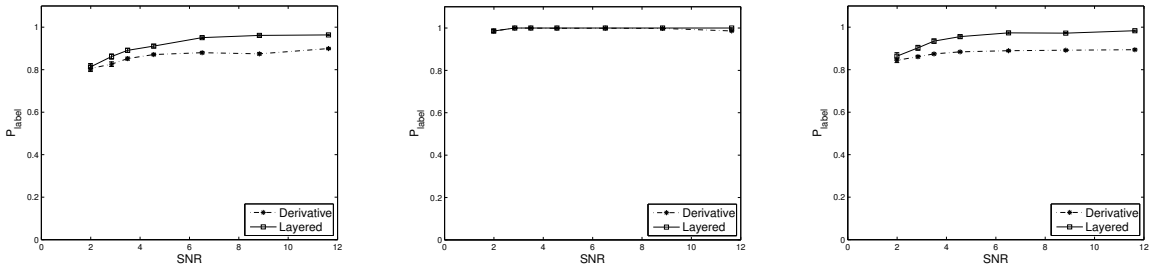
models describe a slight decrease in the intensity over time as well. Except for the sigmoid function, similar models are used for the peak intensity I_{max} in the second channel, and thus no sharp decreases in the intensity are visible in this channel. The models are chosen randomly for each object. If for a certain object the sigmoid model is selected, the corresponding object is assumed to undergo fusion. In this scenario, 12 out of the 30 objects exhibit a fusion behavior. Figure 5.23 displays sample image sections from both channels at a SNR of 2.8. One can observe a significant reduction in the intensity of the objects. We apply the tracking approach to each two-channel image sequence. We use the same parameter values from the first synthetic scenario for the derivative-based approach approach as well the for layered approach.

For each image sequence, we compute the mean labeling accuracy $\bar{P}_{label}$, which is defined as the mean of the labeling accuracy P_{label} obtained over all 30 objects. Then



First channel, $t = 0$ | First channel, $t = 32$ | Second channel, $t = 0$ | Second channel, $t = 32$

Figure 5.23: Second synthetic scenario: Sample time steps of a synthetic image sequence. The SNR is 2.8. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.



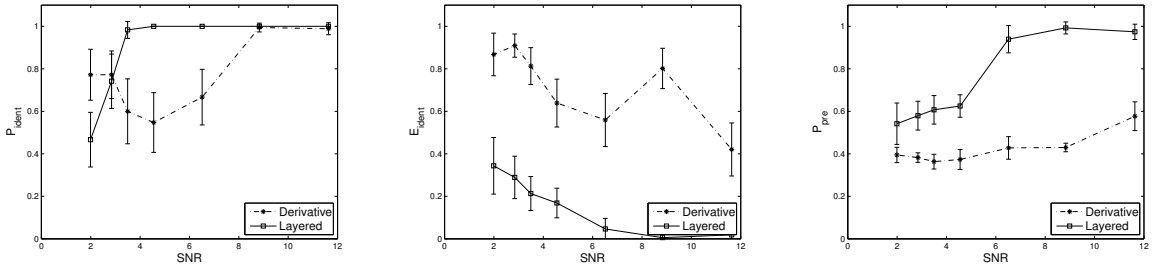
(a) $\bar{P}_{label}$ for the intensity models (first channel)

(b) $\bar{P}_{label}$ for the intensity models (second channel)

(c) $\bar{P}_{label}$ for the behavior

Figure 5.24: Second synthetic scenario: Mean labeling accuracy $\bar{P}_{label}$ as a function of the SNR for the intensity models (first and second channel) as well as for the behavior. The mean values (and standard deviations) of a derivative-based approach ('Derivative') as well as the mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.

we compute the mean values and standard deviations for $\bar{P}_{label}$ over all 30 image sequences for the temporal intensity models in both channels as well as for the behavior. These values are presented in Figure 5.24. Overall for both approaches the labeling performance for the intensity models as well as for the behavior is above 80% for all SNR levels. For example, at a SNR of 2.8 the derivative-based approach achieves a mean value for $\bar{P}_{label}$ of approximately 86% for the behavior while the layered approach attains a mean value for $\bar{P}_{label}$ of 90%. The layered approach outperforms generally the derivative-based approach. The performance degrades smoothly for both approaches as the SNR decreases. In this scenario we also compute the accuracy of behavior identification P_{ident}, the identification error E_{ident}, as well as the precision P_{pre} (see Section 5.2.1) as a function of the SNR. While P_{label} is relatively tolerant to errors in the computed labels for the behavior, P_{ident}, E_{ident}, as well as P_{pre} are relatively sensitive to such errors. As such, a high labeling accuracy P_{label} for the behavior may not necessarily correlate



(a) P_{ident} (b) E_{ident} (c) P_{pre}

Figure 5.25: Second synthetic scenario: Performance in terms of the identification accuracy P_{ident}, identification error E_{ident}, and the precision P_{pre} as a function of the SNR. The mean values (and standard deviations) of a derivative-based approach ('Derivative') as well as the mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.

with a high accuracy of behavior identification P_{ident}, a high precision P_{pre}, or conversely with a low identification error E_{ident}. The mean and standard deviation of these measures computed over 30 image sequences per SNR level are shown in Figure 5.25. The derivative-based approach performs rather poorly in terms of P_{ident}. As the SNR decreases, the identification accuracy decreases, i.e., the ability of the approach to identify particles undergoing fusion is hampered by the noise. In terms of the identification error E_{ident}, the approach yields a large number of false positives, even at high SNR levels. In other words, a large percentage of non-fusion trajectories are mistakenly identified as fusion trajectories. The low precision P_{pre} also indicates that the number of correctly identified fusion trajectories relative to the number of false positives is low, especially at low SNR levels. The layered approach provides a better performance. The accuracy of behavior identification P_{ident} is close to 100% down to a SNR of approximately 3.5. Likewise, the identification error E_{ident} remains below 22% down to a SNR of approximately 3.5. Below a SNR of 3.5 the accuracy of behavior identification P_{ident} decreases rather sharply. Nonetheless, the identification error E_{ident} remains relatively small, which is a favorable property because, while not all true fusion trajectories might be retrieved, the approach also delivers fewer false positives. The precision P_{pre} shows that at large SNR levels the approach identifies fusion trajectories without incurring a relatively large number of false positives. At lower SNR levels the precision P_{pre} decreases relatively strongly due to the increase in the number of false positives. Note that in Figure 5.25a, at low SNR levels, the derivative-based approach achieves a higher P_{ident} compared to the layered approach. This is because at low SNR levels the derivative-based approach (due to its sensitivity to noise) considers most trajectories as fusion trajectories, i.e., the approach correctly identifies most true fusion trajectories (high P_{ident}), but it also

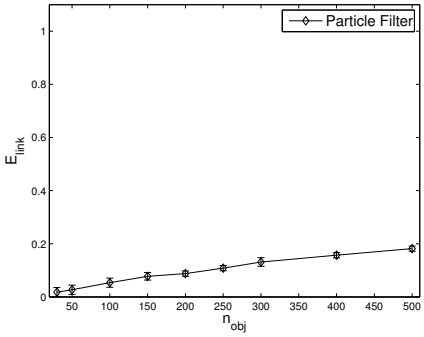
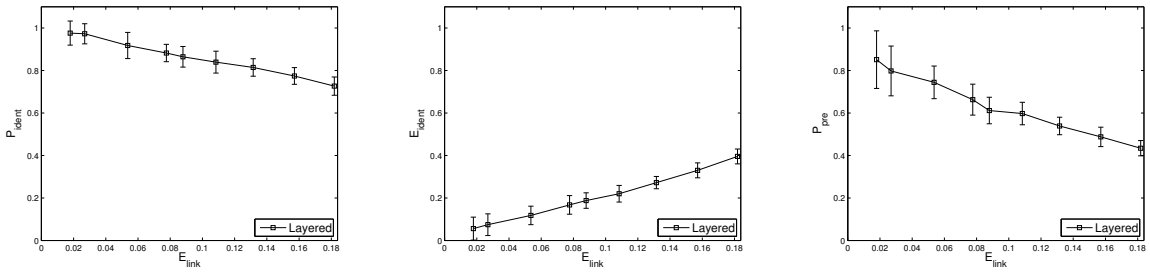


Figure 5.26: Third synthetic scenario: Performance of the tracking approach in terms of the linking error E_{link} as a function of the number of objects N_{obj}. The mean values (and standard deviations) of a tracking approach based on independent particle filters ('Particle Filter') are shown.



(a) P_{ident} (b) E_{ident} (c) P_{pre}

Figure 5.27: Third synthetic scenario: Performance in terms of the identification accuracy P_{ident}, identification error E_{ident}, and the precision P_{pre} as a function of the linking error E_{link}. The mean values (and standard deviations) of the layered probabilistic approach ('Layered') are shown.

considers a large number of non-fusion trajectories as fusion trajectories (high E_{ident}), which is unfavorable. The poorer performance of the derivative-based approach is also reflected by the significantly lower precision P_{pre} compared to the layered approach. In general, the results suggest that the layered approach identifies fusion fairly well given a set of objects exhibiting different behaviors.

5.2.2.3 Third Synthetic Scenario

We also evaluate the performance of the layered approach as a function of the performance of the tracking approach. We generate two-channel image sequences consisting of 50 time steps. The images have dimensions 256×256 pixels (16-bit). To control the performance of the tracking approach, we vary the number of objects in the images using N_{obj} = 30, 50, 100, 150, 200, 250, 300, 400, and 500. The higher N_{obj}, the higher the

object density and thus more tracking errors occur. The initial image position of each object is random and the motion is governed by random walk. The peak intensity I_{max} of each object in each channel is modeled as in the second synthetic scenario (cf. Section 5.2.2.2) and we use SNR = 11.6. For each number of objects N_{obj}, we generate 30 two-channel image sequences (thus in total we use 270 image sequences). We apply our tracking approach based on independent particle filters to each two-channel image sequence. To quantify the tracking performance, we calculate the *linking error* E_{link} [202], [199], which is defined as:

$$E_{link} = 1 - \frac{n_{links,correct}}{n_{links,total}}, \tag{5.7}$$

where $n_{links,correct}$ is the number of correct links and $n_{links,total}$ is the number of true links. A link corresponds to a displacement vector between two consecutive positions of a trajectory. A correct link corresponds to a displacement vector between two positions close to two consecutive positions of a true trajectory. We apply the layered approach for behavior identification using the computed trajectories. We use the same parameter values for the layered approach as in the first and second synthetic scenarios.

For each number of objects N_{obj}, we compute the mean value for E_{link} over the 30 image sequences (see Figure 5.26). It can be seen that E_{link} increases linearly with N_{obj} and that the tracking approach copes relatively well with a large number of objects (e.g., for $N_{obj} = 500$, $E_{link} = 18\%$). We also determine the accuracy of behavior identification P_{ident}, the identification error E_{ident}, as well as the precision P_{pre} as a function of E_{link}. The result in Figure 5.27a shows that P_{ident} degrades slowly as E_{link} increases. In Figure 5.27b, E_{ident} exhibits a linear relation with E_{link} which suggests that the number of false positives is directly influenced by tracking errors. The precision P_{pre} (Figure 5.27c) decreases as E_{link} increases. To summarize, the layered approach copes well with a small number of tracking errors. For a larger number of tracking errors, the performance degrades, however, the degradation is smooth.

5.2.3 Evaluation on Real Microscopy Images

We have also applied the layered approach to real microscopy image sequences displaying HIV-1 particles. We have considered two scenarios each involving a different technique for fluorescently labeling the virus particles yielding different kinds of image data. We have carried out an experimental comparison with a derivative-based approach [61]. In addition, we perform a comparison with a trajectory-based approach (cf. [70]) that compares the endpoints of trajectories obtained by tracking using a single channel with the endpoints of trajectories obtained by tracking using two channels. A fusion event

Table 5.5: First real scenario: Description of the real image sequences.

Sequence	No. of evaluated trajectories	No. of fusion trajectories
1	20	1
2	20	0
3	20	0
4	20	0
5	20	0

is identified when the endpoint of a single-channel trajectory differs from the endpoint of a corresponding two-channel trajectory. Since this approach based on trajectory endpoints only takes into consideration spatial information, it cannot determine the occurring temporal intensity models of an individual particle. Thus P_{label} for the intensity models cannot be calculated. Likewise, since the behaviors are defined in terms of the intensity models (see Section 4.5.2), the occurring behavior for each time step could not be determined and so P_{label} for the behavior is not computed. Since the approach based on trajectory endpoints categorizes each trajectory as a fusion or non-fusion trajectory, we report results for P_{ident}, E_{ident}, and P_{pre}.

5.2.3.1 First Real Scenario

In the first real scenario, the outer shell (viral matrix) of individual pseudotyped HIV-1 particles is labeled with an enhanced green fluorescent protein (MA.eGFP) while the inner core part (viral protein R) is tagged with the red fluorescent protein (mRFP.Vpr). The HIV-1 particles are pseudotyped with the glycoprotein of the vesicular stomatitis virus (VSV-G) and incubated with HeLa cells [108]. Images are acquired using a Zeiss Axiovert 200 M microscope with a Roper Scientific Cascade II EM-CCD. A pair of images (one per channel) is recorded every 100 ms. Upon fusion, the label attached to the outer shell dissolves, and thus a decrease in the intensity in the corresponding channel is observed. The acquired image sequences consist of 200-400 two-channel images (512×512 pixels; 16-bit). Within this scenario, we evaluate 5 image sequences. To track the virus particles in the two-channel image sequences, we use Gaussian fitting and independent particle filters [70]. Ground truth for the intensity labels is obtained manually for 20 of the computed trajectories within each image sequence by inspecting the intensity over time of the tracked virus particles. The labels for the behavior are computed via the composite function $(h^{-1} \circ g)(\boldsymbol{\alpha})$ on the manually determined intensity models $\boldsymbol{\alpha}$. Based on the labels for the behavior, trajectories are categorized as fusion

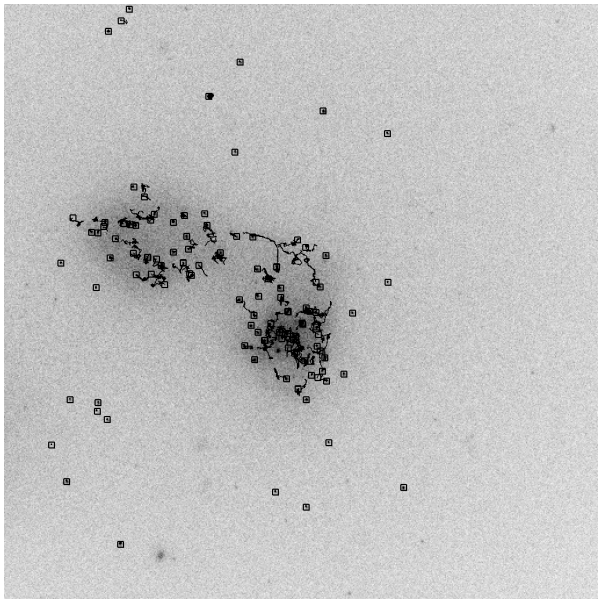


Figure 5.28: First real scenario: Tracking results for the real image sequence "Sequence 1". The time step is $t = 199$. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

or non-fusion trajectories; see Table 5.5. For the evaluation, the parameter values for the derivative-based approach as well as for the layered approach are kept fixed for all image sequences.

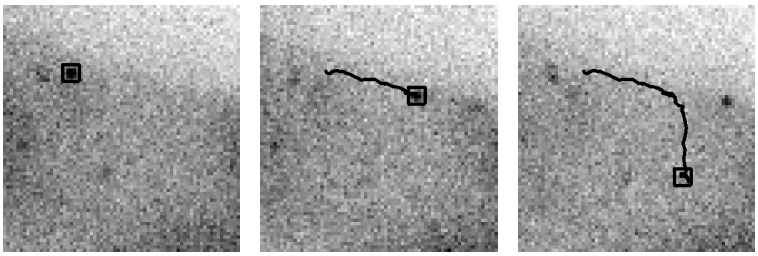
Images from Sequence 1 in Table 5.5 and the corresponding tracking results are shown in Figure 5.28. A particle undergoing fusion is displayed in Figure 5.29. The ground truth labels for the intensity models for either channel reflect a decrease in the intensity observed in the first channel that corresponds to fusion (see Figure 5.30a) as well as an increase in the intensity in the second channel (see Figure 5.30b). The corresponding changes in the intensity are illustrated in Figure 5.31a and Figure 5.31d, where the measured intensity over time computed for each channel is shown. The computed derivative values for the first channel as well as for the second channel are shown in Figure 5.31b and Figure 5.31e, respectively. The results for the derivative-based approach for the estimated intensity models for the first and second channels are shown in Figure 5.31c and Figure 5.31f. Qualitatively, the derivative-based approach does not recover very well the motif underlying the true sequence of the intensity models. In particular, wrong labels scattered throughout the computed sequences of the intensity models lead to fragmented segments within these sequences. Relative to the ground truth labels, the derivative-based approach achieves a labeling accuracy of $P_{label} = 92\%$ for the intensity models of the first channel and $P_{label} = 94\%$ for the intensity models of the second channel. In comparison, the layered approach retrieves the underlying

Table 5.6: First real scenario: Results for real image sequences in terms of the mean labeling accuracy $\bar{P}_{label}$ [%] for the temporal intensity models in both channels as well as for the behavior. The mean values and standard deviations are also shown.

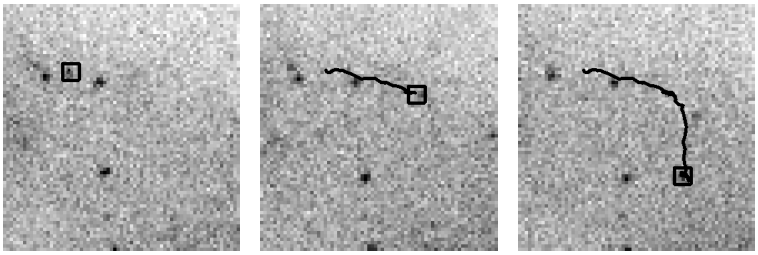
	Derivative			Layered		
Sequence	1st Ch.	2nd Ch.	Behavior	1st Ch.	2nd Ch.	Behavior
1	96	94	87	98	97	93
2	99	96	95	100	98	98
3	99	89	85	100	97	93
4	94	88	82	96	91	86
5	97	92	91	98	95	95
Mean	97	92	88	98	96	93
Std. Dev.	2	3	5	1	3	5

motif relatively well (cf. Figure 5.32c and Figure 5.32f). Note the timely activation of the appropriate intensity models (cf. Figure 5.32b and Figure 5.32e). Accordingly, the approach achieves a labeling accuracy of $P_{label} = 96\%$ for the first channel and $P_{label} = 95\%$ for the second channel. Figure 5.33 displays the ground truth labels for the behavior as well as the labels for the behavior computed by the derivative-based approach as well as by the layered approach. Both approaches identify the fusion behavior. For the derivative-based approach, the fragmentation observed in the sequences of intensity models carries over to the sequence of behaviors. This leads to a lower labeling accuracy for the behavior ($P_{label} = 87\%$). In contrast, the layered approach yields fairly accurate results, as reflected by a higher labeling accuracy for the behavior ($P_{label} = 95\%$).

Table 5.6 displays the performance of the derivative-based scheme as well as the layered approach in terms of the mean labeling accuracy $\bar{P}_{label}$ for the intensity models on both channels as well as for the behavior. The mean and standard deviation over all five two-channel image sequences are shown as well. Overall, the labeling accuracy for the intensity models is quite high for both approaches (the mean value for $\bar{P}_{label}$ is above 90%). For the behavior, the derivative-based scheme yields a mean value for $\bar{P}_{label}$ of 88% (standard deviation of 5%) while the layered approach delivers a mean value for $\bar{P}_{label}$ of 93% (standard deviation of 5%). In all sequences, the layered approach achieves a better labeling performance than the derivative-based approach. The accuracy of behavior identification P_{ident}, the identification error E_{ident}, as well as the precision P_{ident} over all sequences are shown in Table 5.7. For instance, for Sequence 1

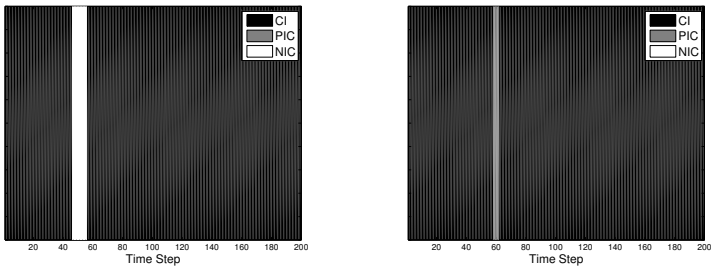


First channel, $t = 0$ First channel, $t = 40$ First channel, $t = 80$



Second channel, $t = 0$ Second channel, $t = 40$ Second channel, $t = 80$

Figure 5.29: First real scenario: Tracking results for a virus particle undergoing fusion. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

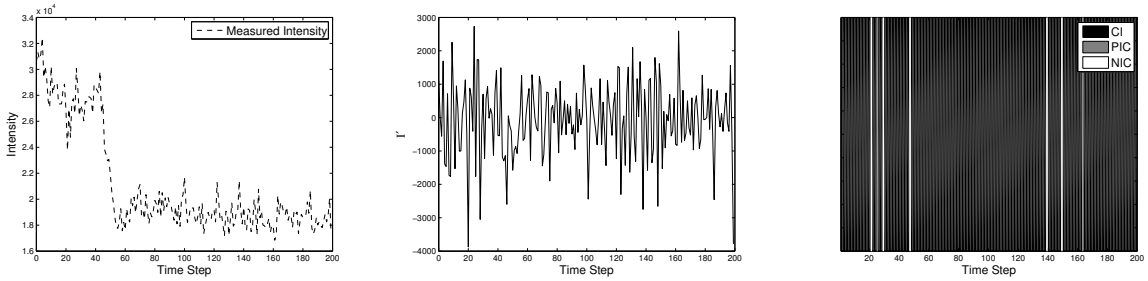


(a) True labels for the intensity models (first channel)

(b) True labels for the intensity models (second channel)

Figure 5.30: First real scenario: Ground truth for the intensity models.

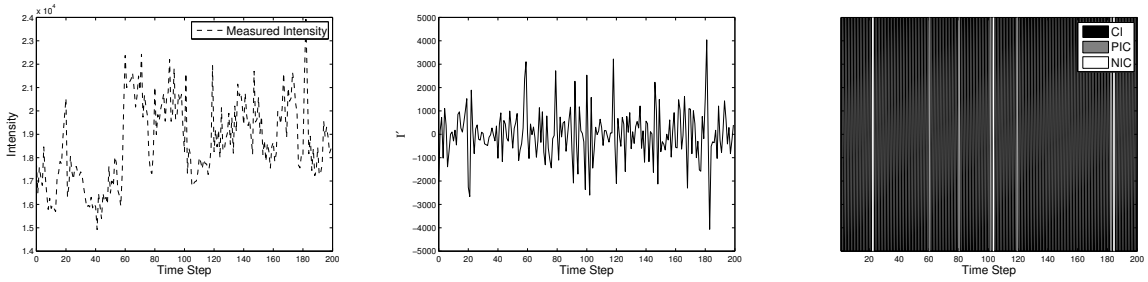
we obtain $P_{ident} = 100\%$ for the derivative-based approach, for the approach based on trajectory endpoints, as well as for the layered approach. The derivative-based approach yields an identification error of $E_{ident} = 58\%$ while the approach based on trajectory endpoints incurs $E_{ident} = 5\%$. The reason for the high value of E_{ident} for the derivative-based approach is that this approach is quite sensitive to image noise. The approach based on trajectory endpoints in comparison does not directly evaluate the image intensities. However, this approach also delivers a certain number of false positives. In comparison, the layered approach achieves $E_{ident} = 0\%$. The precision of the derivative-based approach ($P_{pre} = 8\%$) also reflects the larger number of false positives generated by this approach. The approach based on trajectory endpoints yields $P_{pre} = 50\%$. A higher precision is achieved by the layered approach ($P_{pre} = 100\%$). The other image



(a) Measured intensity (first channel)

(b) Computed derivative of the measured intensity (first channel)

(c) Computed labels for the intensity models (first channel)



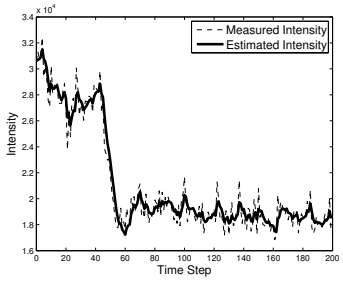
(d) Measured intensity (second channel)

(e) Computed derivative of the measured intensity (second channel)

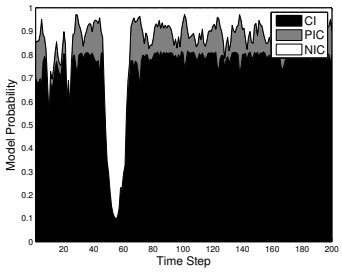
(f) Computed labels for the intensity models (second channel)

Figure 5.31: First real scenario: Results obtained using a derivative-based approach. The labeling accuracy is $P_{label} = 92\%$ for the first channel and $P_{label} = 94\%$ for the second channel.

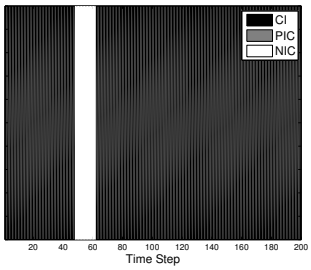
sequences studied within this scenario do not display particles undergoing fusion and therefore P_{ident} is not calculated. For the identification error E_{ident}, the derivative-based approach yields a mean value of $\bar{E}_{ident} = 37\%$ (standard deviation of 13%). The values for E_{ident} of the derivative-based approach are thus in accordance with the results obtained for the second synthetic scenario (cf. Figure 5.25), where the derivative-based approach yielded a large identification error E_{ident}. In comparison, the approach based on trajectory endpoints achieves a mean value of $\bar{E}_{ident} = 25\%$ (standard deviation of 15%). This suggests that this approach yields a lower number of false positives compared to the derivative-based approach. The layered approach achieves the lowest mean value for the identification error ($\bar{E}_{ident} = 0\%$, standard deviation of 0%)) and the result is in agreement with the results obtained for the synthetic images (cf. Figure 5.25), where E_{ident} of the layered approach is significantly lower compared to the derivative-based approach. Because some image image sequences do not display fusion particles, and because in those cases the derivative-based approach as well as the approach based on trajectory endpoints yield a certain number of false positives, the precision P_{pre} of these approaches is zero in these cases. Because the layered approach does not yield false positives, P_{pre} cannot be calculated (division by zero). A more comprehensive analysis of the performance in terms of P_{ident} as well as P_{pre} requires a larger number of image



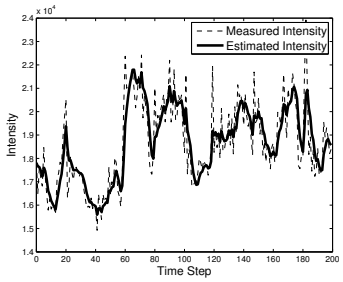
(a) Measured and estimated intensity (first channel)



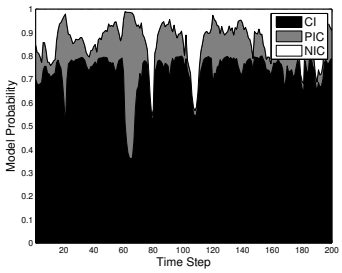
(b) Computed posterior model probabilities for the intensity models (first channel)



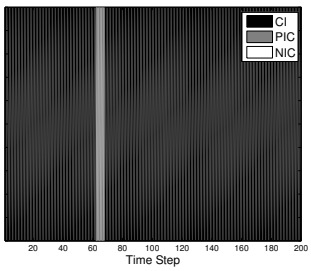
(c) Computed labels for the intensity models (first channel)



(d) Measured and estimated intensity (second channel)

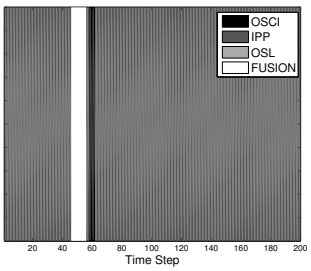


(e) Computed posterior model probabilities for the intensity models (second channel)

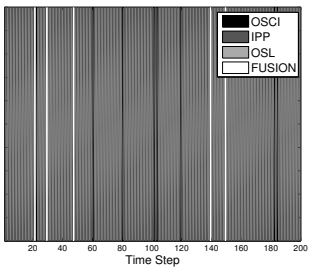


(f) Computed labels for the intensity models (second channel)

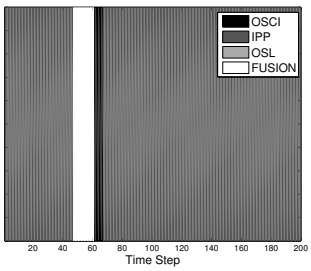
Figure 5.32: First real scenario: Results obtained using a hybrid particle filter. The labeling accuracy is $P_{label} = 96\%$ for the first channel and $P_{label} = 95\%$ for the second channel.



(a) True labels for the behavior



(b) Computed labels for the behavior using a derivative-based approach



(c) Computed labels for the behavior using the layered approach

Figure 5.33: First real scenario: Ground truth for the behavior and results obtained using a derivative-based approach as well as the layered probabilistic approach. The labeling accuracy for the derivative-based approach is $P_{label} = 87\%$ while for the layered approach the labeling accuracy is $P_{label} = 95\%$.

Table 5.7: First real scenario: Results for real image sequences in terms of the accuracy for behavior identification P_{ident} [%], the identification error E_{ident} [%], and the precision P_{pre} [%]. The mean values and standard deviations are also shown. Since in certain image sequences no fusion events occurred, the corresponding P_{ident} values could not be computed (–). In cases where no fusion events occurred nor false positives were obtained, the P_{pre} values could not be computed.

	Derivative			Trajectory Endpoints			Layered		
Sequence	P_{ident}	E_{ident}	P_{pre}	P_{ident}	E_{ident}	P_{pre}	P_{ident}	E_{ident}	P_{pre}
1	100	58	8	100	5	50	100	0	100
2	–	35	0	–	20	0	–	0	–
3	–	30	0	–	40	0	–	0	–
4	–	35	0	–	40	0	–	0	–
5	–	25	0	–	20	0	–	0	–
Mean	–	37	2	–	25	10	–	0	–
Std. Dev.	–	13	4	–	15	22	–	0	–

sequences displaying particles undergoing fusion. Since we observed very few fusion events using the fluorescent labeling technique (MA.eGFP, mRFP.Vpr) [108] within this scenario, an improved labeling technique was developed that led to a larger number of fusion occurrences.

5.2.3.2 Second Real Scenario

In the second real scenario, the outer shell (lipid envelope) of HIV-1 particles pseudotyped with the Env glycoprotein of the ecotropic murine leukemia virus is labeled with the yellow fluorescent protein (MLVEnv.YFP). The mCherry fluorescent protein (MA.mCherry) is attached to the inner part (viral matrix) [103]. Images are acquired using the same wide-field microscopy setup (but in TIRF mode) as in the first real scenario. Here, the acquisition rate for a pair of images is 2 seconds, which experimentally was found to be sufficient. Fusion entails a loss of the label attached to the outer shell (lipid envelope) along with a decrease in intensity in the corresponding channel. However, a decrease in intensity in *both* channels is also indicative of fusion when using this labeling strategy. This variant is accommodated into the evaluated approaches by adjusting the function $g(\boldsymbol{\alpha})$. We evaluate 8 image sequences. We use our approach based on Kalman filters to track the virus particles (cf. [70]). The ground truth for the fusion events is determined manually. A summary of the real image data is shown in

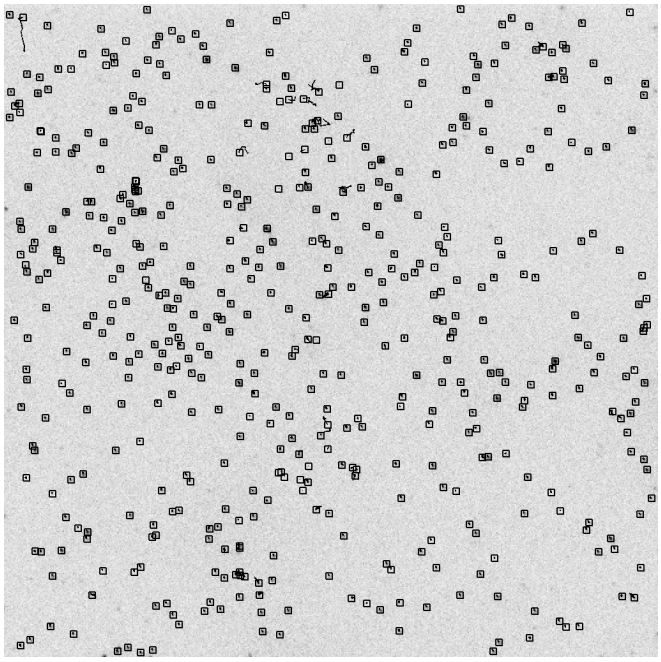


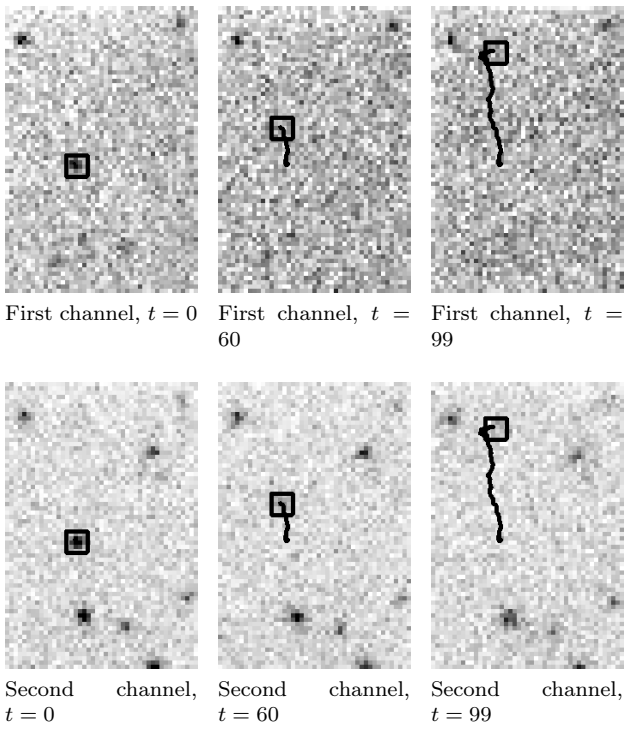
Figure 5.34: Second real scenario: Tracking results for the real image sequence "Sequence 2". The time step is $t = 99$. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

Table 5.8.

As an example, images from Sequence 2 in Table 5.8 and the respective tracking results are shown in Figure 5.34. Notwithstanding the relatively high object density, the tracking approach determines successfully the trajectories of the virus particles. A virus particle undergoing fusion is shown in Figure 5.35. In this case, the contrast of the first channel is much lower, which entails subtler changes in the intensity. As fusion occurs, the intensity drops almost simultaneously in both channels. Accordingly, the ground truth labels for the intensity models for both channels include concurrent time periods where the 'Negative Intensity Change' (NIC) model is dominant (see Figure 5.36). The measured intensity for both channels shown in Figure 5.37a and Figure 5.37d reflects more explicitly the aforementioned drop in intensity. Here the computed derivative values are very noisy (Figure 5.37b and Figure 5.37e) and consequently the derivative-based approach yields fragmented sequences for the estimated intensity models (Figure 5.37c and Figure 5.37f). The labeling accuracy for the intensity models is $P_{label} = 75\%$ for the first channel and $P_{label} = 91\%$ for the second channel. The results obtained by the layered approach are displayed in Figure 5.38. The approach achieves $P_{label} = 88\%$ for the first channel and $P_{label} = 85\%$ for the second channel. Thus, the result for the first channel is much better. The reason for the lower performance for the second channel is that the hybrid particle filter did not adapt quickly enough to the fusion-induced change.

Table 5.8: Second real scenario: Description of the real image sequences.

Sequence	No. of evaluated trajectories	No. of fusion trajectories
1	20	2
2	20	1
3	20	2
4	20	1
5	20	1
6	20	1
7	20	2
8	20	1

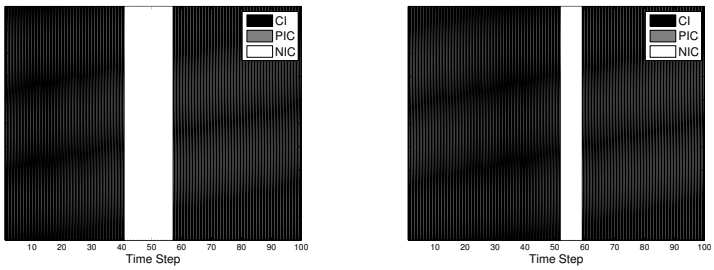


First channel, $t = 0$ | First channel, $t = 60$ | First channel, $t = 99$

Second channel, $t = 0$ | Second channel, $t = 60$ | Second channel, $t = 99$

Figure 5.35: Second real scenario: Tracking results for a virus particle undergoing fusion. For visualization purposes, the image contrast has been enhanced and the image intensities have been inverted.

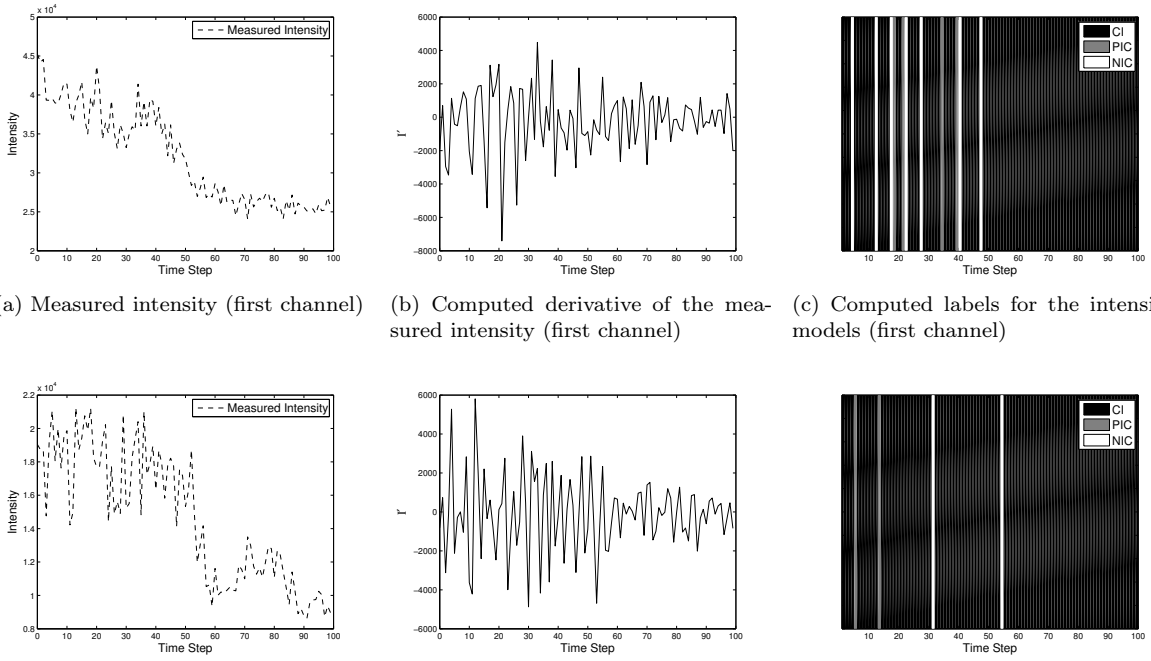
This could be addressed by choosing different values for the parameters (e.g., the noise parameter σ_n). However, the adaptation time and the steady behavior of the hybrid particle filter need to be balanced. The ground truth labels for the behavior of this particle are shown in Figure 5.39. For the derivative-based approach erroneous labels for the intensity models computed on either channel are propagated to the labels for the



(a) True labels for the intensity models (first channel)

(b) True labels for the intensity models (second channel)

Figure 5.36: Second real scenario: Ground truth for the intensity models.



(a) Measured intensity (first channel)

(b) Computed derivative of the measured intensity (first channel)

(c) Computed labels for the intensity models (first channel)

(d) Measured intensity (second channel)

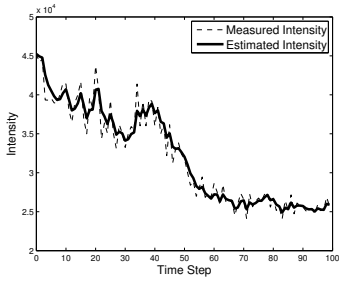
(e) Computed derivative of the measured intensity (second channel)

(f) Computed labels for the intensity models (second channel)

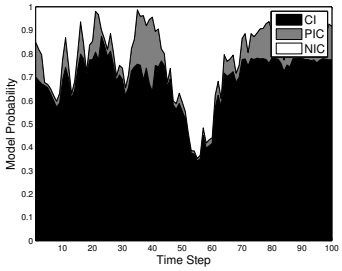
Figure 5.37: Second real scenario: Results obtained using a derivative-based approach. The labeling accuracy is $P_{label} = 75\%$ for the first channel and $P_{label} = 91\%$ for the second channel.

behavior. Accordingly, the labeling performance for the behavior is $P_{label} = 74\%$. For the layered approach, the computed behaviors are in better agreement with the ground truth labels yielding a significantly higher labeling accuracy of $P_{label} = 87\%$.

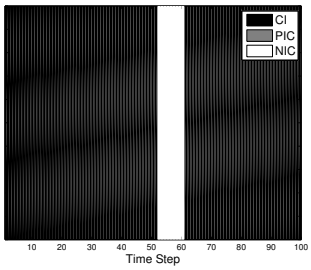
As shown in Table 5.9, the overall labeling accuracy for the derivative-based approach in this scenario is worse compared to the first real scenario (see Table 5.6). The reason for this is that in the second real scenario the changes in the intensity corresponding to fusion are smaller. To detect such weaker changes, the threshold applied to the derivative values has to be lowered. Because of the noise, the low threshold leads to a larger number



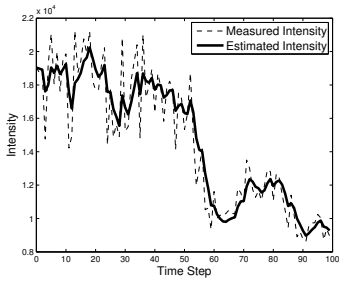
(a) Measured and estimated intensity (first channel)



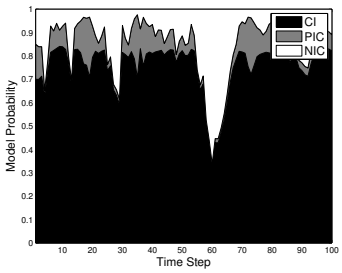
(b) Computed posterior model probabilities for the intensity models (first channel)



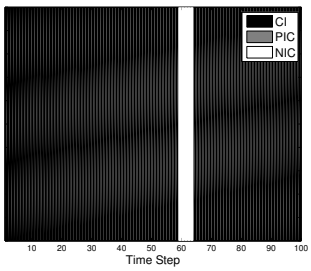
(c) Computed labels for the intensity models (first channel)



(d) Measured and estimated intensity (second channel)

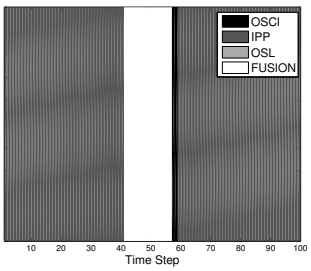


(e) Computed posterior model probabilities for the intensity models (second channel)

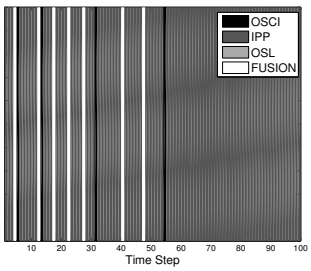


(f) Computed labels for the intensity models (second channel)

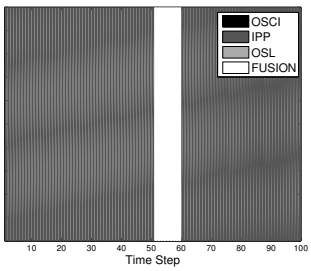
Figure 5.38: Second real scenario: Results obtained using a hybrid particle filter. The labeling accuracy is $P_{label} = 88\%$ for the first channel and $P_{label} = 85\%$ for the second channel.



(a) True labels for the behavior



(b) Computed labels for the behavior using a derivative-based approach



(c) Computed labels for the behavior using the layered approach

Figure 5.39: Second real scenario: Ground truth for the behavior and results obtained using a derivative-based approach as well as the layered probabilistic approach. The labeling accuracy for the derivative-based approach is $P_{label} = 74\%$ while for the layered approach the labeling accuracy is $P_{label} = 87\%$.

Table 5.9: Second real scenario: Results for real image sequences in terms of the mean labeling accuracy $\bar{P}_{label}$ [%] for the temporal intensity models in both channels as well as for the behavior. The mean values and standard deviations are also shown.

	Derivative			Layered		
Sequence	1st Ch.	2nd Ch.	Behavior	1st Ch.	2nd Ch.	Behavior
1	95	86	90	97	97	98
2	90	90	88	98	98	99
3	66	46	56	96	90	95
4	98	98	97	98	99	99
5	86	99	92	97	99	98
6	100	100	100	100	100	100
7	88	98	92	89	98	98
8	82	62	71	96	96	98
Mean	88	85	86	96	97	98
Std. Dev.	11	20	15	3	3	1

Table 5.10: Second real scenario: Results for real image sequences obtained in terms of the accuracy for behavior identification P_{ident} [%], the identification error E_{ident} [%], and the precision P_{pre} [%]. The mean values and standard deviations are shown, too.

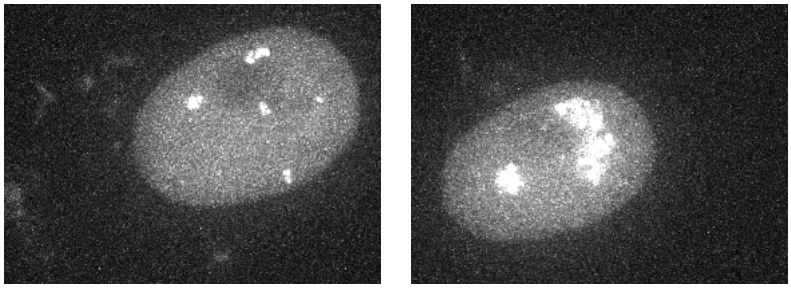
	Derivative			Trajectory Endpoints			Layered		
Sequence	P_{ident}	E_{ident}	P_{pre}	P_{ident}	E_{ident}	P_{pre}	P_{ident}	E_{ident}	P_{pre}
1	50	83	6	50	22	20	50	0	100
2	100	100	5	100	16	25	100	0	100
3	100	100	10	50	17	25	100	0	100
4	100	26	17	0	0	–	100	5	50
5	100	95	5	100	11	33	100	11	33
6	100	11	33	100	32	14	100	0	100
7	100	83	12	0	6	0	100	17	40
8	100	100	5	100	42	11	100	11	33
Mean	94	75	12	62	18	18	94	5	70
Std. Dev.	18	36	10	44	14	11	18	7	33

of errors for the labels of the intensity models. The errors in the labels for the intensity models are propagated to the labels for the behavior, and thus the performance of this approach suffers. Concretely, the mean value of $\bar{P}_{label}$ for the intensity models as well as for the behaviors is below 90%. The layered approach instead reconstructs the underlying intensity via the proposed intensity models. A good reconstruction enhances the dynamic properties of the underlying signal, which leads to a correct identification of the dominant intensity models. Correct identification of the intensity models in turn provides improved predictions for the intensity at subsequent time steps. This indicates that the estimates for the intensity and the intensity models benefit from the joint estimation process embodied by the stochastic hybrid system. For the layered approach, the mean value of $\bar{P}_{label}$ for the intensity models as well as for the behaviors is above 95%. In Table 5.10, the performance of the approaches (and that of the approach based on trajectory endpoints) is further reflected by the accuracy of behavior identification P_{ident}, the identification error E_{ident}, as well as the precision P_{pre} computed for all eight image sequences. Here, the derivative-based approach as well as the layered approach achieve a mean accuracy of behavior identification of $\bar{P}_{ident} = 94\%$ (standard deviation of 18% for both approaches). That is, both approaches recover well the true fusion trajectories. In comparison, the approach based on trajectory endpoints yields $\bar{P}_{ident} = 62\%$ (standard deviation of 44%). The reason for this is that this approach assumes that fusion is described by a decrease in intensity in a single channel only. In this scenario, however, fusion is described by a decrease in intensity in *both* channels.

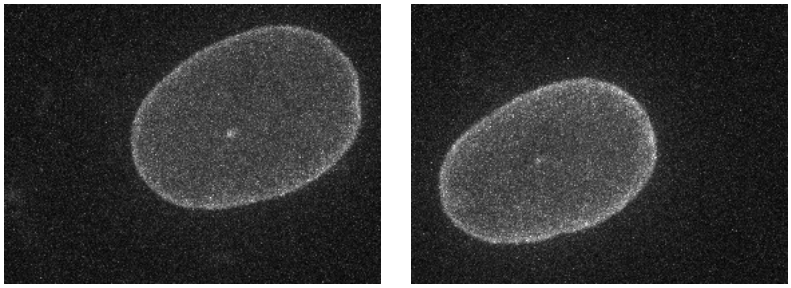
For the identification error E_{ident}, the derivative-based approach delivers a mean value of $\bar{E}_{ident} = 75\%$ (standard deviation of 36%), which further confirms the sensitivity of the approach to the image noise. The result for the approach based on trajectory endpoints is moderate ($\bar{E}_{ident} = 18\%$, standard deviation of 14%). The best result is achieved by the layered approach ($\bar{E}_{ident} = 5\%$, standard deviation of 7%).

In terms of the precision P_{pre}, the derivative-based approach achieves a mean value of $\bar{P}_{pre} = 12\%$ (standard deviation of 10%), where the low precision may be attributed to the large number of false positives. The approach based on trajectory endpoints also yields a low precision ($\bar{P}_{pre} = 18\%$, standard deviation of 11%), which can be explained by the low number of correctly identified fusion trajectories. The layered approach achieves $\bar{P}_{pre} = 70\%$ (standard deviation of 33%), which highlights the improved performance compared to previous approaches. Overall, the results suggest that the layered approach is well suited for the task of identifying fusion of HIV-1 particles.

The approach was implemented in Java within our software ViroTracker ([70]). The computation time of the approach scales linearly with respect to the number of samples for the hybrid particle filter and quadratically with respect to the number of states in



GFP, $t = 0$ GFP, $t = 124$



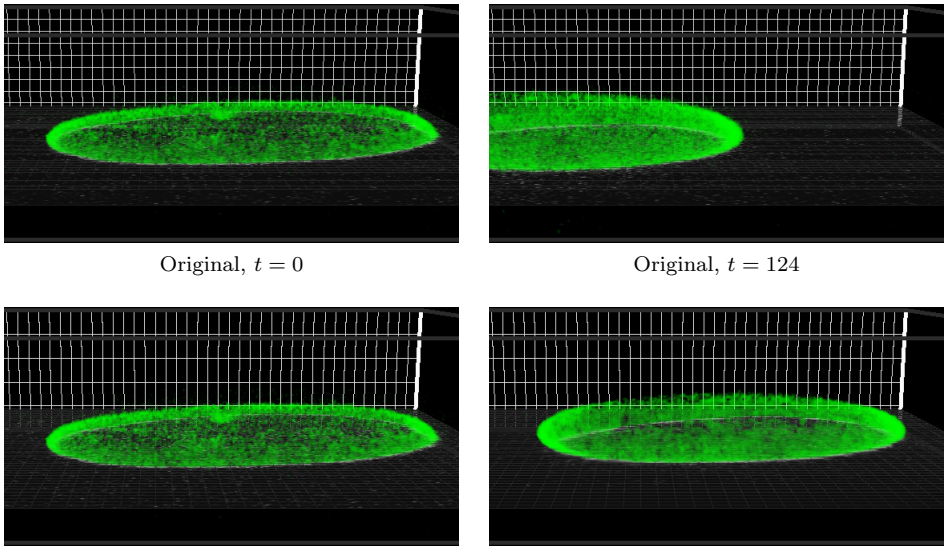
mCherry-LA, $t = 0$ mCherry-LA, $t = 124$

Figure 5.40: Real microscopy images displaying HSV replication compartments (GFP channel) inside a nucleus (mCherry-LA channel). Maximum intensity projections are shown.

the top-most Markov chain (i.e., number of behaviors). For example, the computation time for one trajectory comprising 100 time steps and using 1000 samples is ca. 1 second on an AMD Opteron (2.3 GHz) CPU running Linux.

5.3 Tracking and Motion Analysis of HSV Compartments

We have also applied our tracking approaches to 3D image sequences displaying herpes simplex virus (HSV) replication compartments [35]. Viral chromosomes are replicated in intranuclear structures called replication compartments. Studying the motion of the replication compartments reveals the reproduction strategies of the viruses. In particular, the mechanisms and functions underlying the motion of HSV compartments are not well known. To observe the motion of HSV compartments, the structures are tagged with the GFP label while the nuclei within which the compartments are located are tagged with the mCherry-LA label. A Zeiss Axiovert 200M microscope was used to acquire a pair of image stacks (one stack corresponding to the compartments and one stack corresponding to the nucleus) every minute over an observation period of up 3 hours. At each time step, stacks of up 32 images (z-slices) with dimensions of up to 512×512 pixels (16-bit) were acquired. Sample images from both channels are shown



Original, $t = 0$ | Original, $t = 124$

Registered, $t = 0$ | Registered, $t = 124$

Figure 5.41: Original images and registration results. The nuclei are volume rendered in green.

in Figure 5.40. In total, we analyzed 26 image sequences.

The replication compartments are located within the nucleus. Since the nucleus undergoes motion, the observed motion of the compartments is a superposition of the nucleus motion and the compartments' motion. Figure 5.40 displays an example of a nucleus undergoing a large translational motion. To remove the translational and rotational motion of the observed nuclei, we developed a rigid registration approach. With this approach, the following steps were carried out. For each time step of an image sequence, we detected the image region corresponding to the nucleus using a 3D Gaussian filter and an intensity threshold. The values for the standard deviations $\sigma_{\mathrm{F},xy}$ and $\sigma_{\mathrm{F},z}$ of the Gaussian filter were chosen based on the size of the nucleus (typically, we used values of $\sigma_{\mathrm{F},xy} = 10$ voxels and $\sigma_{\mathrm{F},z} = 2$ voxels). The position of the nucleus was determined by computing the center of mass of the corresponding voxel positions, whereas the orientation was obtained by diagonalizing the covariance matrix of the voxel positions. Each time step t of the image sequence was aligned to time step zero by computing a rigid transformation based on the position and orientation of the nucleus at both time steps. Subsequently, we applied an intensity-based rigid registration scheme [217]. Because of photobleaching, the intensities of the first time step differed strongly from those of the latter time steps. To cope with this, we applied the registration approach to consecutive time steps of the image sequence, thereby obtaining a rigid transformation between time steps t and $t + 1$. By recursively concatenating the sequential transformations, we computed the required transformation between time steps zero and t. We applied this registration approach to the images in the mCherry

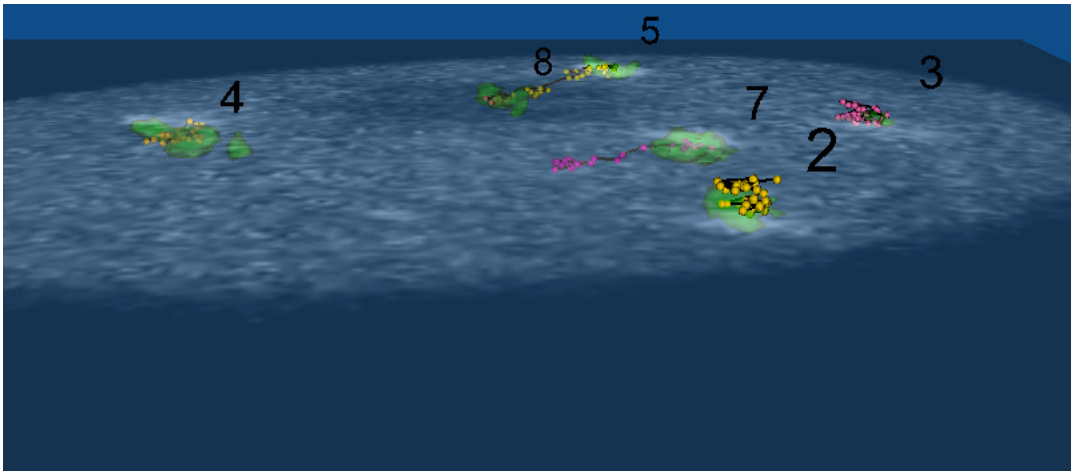


Figure 5.42: Trajectories of HSV compartments as determined by our tracking approach. Trajectories are rendered as spheres (positions) and sticks (displacement vectors). The numbers identify each trajectory. The compartments are represented via the green surfaces. A z-slice of the original images ($z = 21$) is shown. The time step is $t = 31$.

channel. The resulting transformations were also applied to the corresponding images in the GFP channel. Sample registration results for the images shown Figure 5.40 are displayed in Figure 5.41. We used the registered 3D image sequences as input for the tracking approach. Note that the appearance of the replication compartments does not resemble well a Gaussian function. The tracking approaches based on the particle filter as well as on the PDAE approach assume that the appearance of the objects is well described by a Gaussian function. Since the appearance of the compartments does not conform well to a Gaussian function, we used an approach based on the spot-enhancing filter (see Section 3.1) for localization and the Kalman filter (see Section 3.5 as well as Table 5.3) with a Brownian motion model for estimating the position over time of the compartments. Tracking results for the registered images shown in Figure 5.41 are presented in Figure 5.42. It can be seen that the tracking approach yields good results.

To determine the motion type of individual compartments, we developed a hierarchical approach based on both the anomalous diffusion coefficient α [165] as well as the 3D relative anisotropy κ^2 [189], [164]. To obtain an estimate for the anomalous diffusion coefficient α of a trajectory, we computed the mean-square displacement (MSD) as a function of the time interval Δt. We fit the anomalous diffusion model $\text{MSD} = 6D\Delta t^{\alpha}$ [165] to the calculated MSD values, thereby obtaining an estimate for the diffusion coefficient D as well as for the anomalous diffusion coefficient α. To improve the accuracy of the estimates, we restricted the MSD calculations to time intervals t smaller than $N_{\text{steps}}/3$, where N_{steps} is the total number of available positions in the trajectory. To characterize the shape of the trajectory by its 3D relative anisotropy κ^2, we first calcu-

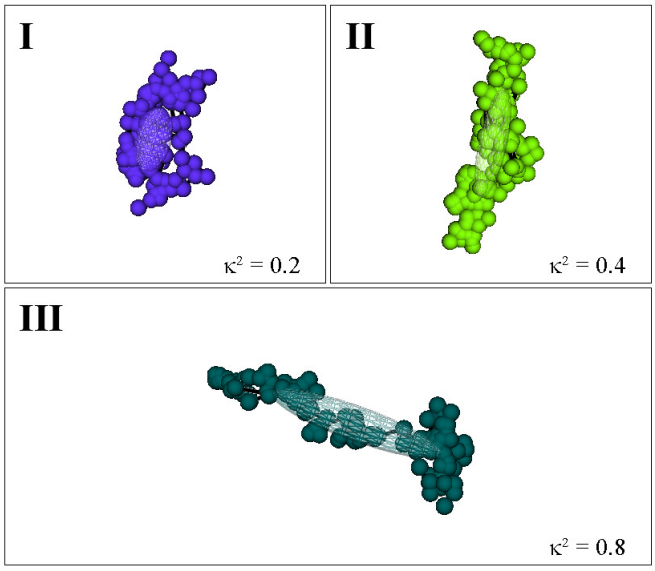


Figure 5.43: Sample 3D trajectories and their ellipsoids of gyration are shown along with their κ^2 values, which are defined as a function of the squared length of the semiaxes of the ellipsoid. Trajectories I and II represent non-directed motion, and trajectory III represents directed motion. Corresponding MSD curves are shown in Figure 5.44.

lated the squared radius of gyration s^2, the asphericity b, and the acylindricity c using the eigenvalues of the covariance matrix of the trajectory's positions. Based on these shape measures, we calculated the relative shape anisotropy:

$$\kappa^2 = \frac{b^2 + 0.75c^2}{s^4}. \tag{5.8}$$

Note that $\kappa^2 \in [0, 1]$, where values close to 0 correspond to trajectories with an isotropic shape (i.e., trajectories exhibiting random motion) while values tending towards 1 correspond to trajectories with an anisotropic shape (i.e., trajectories displaying directed motion). Thus, we used κ^2 as our criterion to detect directed motion using a threshold value of $\kappa^2 = 0.5$. This threshold value was determined using simulated trajectories where the true motion type of the trajectories was known. Our hierarchical approach for determining the motion type of the compartments, therefore, first used κ^2 to distinguish trajectories exhibiting directed motion from those displaying random motion. We then used α to classify the latter into confined diffusion ($\alpha < 0.1$), obstructed diffusion ($0.1 \leq \alpha < 0.9$), or simple diffusion ($\alpha \geq 0.9$) (see also Bacher et al. [10]). Sample trajectories of HSV compartments and an illustration of the motion parameters κ^2 as well as α are shown in Figures 5.43 and 5.44. A two-sample Kolmogorov-Smirnov test was used to determine the statistical significance of differences between different experimental conditions in the cumulative distribution of κ^2 values.

We analyzed the behavior of the compartments under different experimental condi-

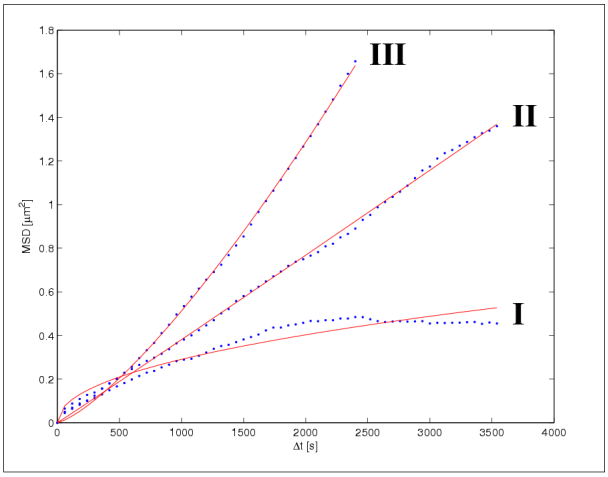


Figure 5.44: Corresponding MSD curves (blue) and fitted curves (red) of the trajectories in Figure 5.43 are shown. Based on their α-values, trajectories I and II are further classified as obstructed diffusion and simple diffusion, respectively.

tions. Under the control (wildtype, WT) conditions, we determined the position over time of 23 compartments and applied our hierarchical motion classification scheme. It turned out that the majority (74%) of replication compartments undergo directed motion (see WT in Figure 5.45). No compartment displayed confined diffusion. Thus the replication compartments move via an active process. To determine which nuclear structures are involved in the motion process, cells were treated with the Butanedione monoxime (BDM) drug that impairs the function of the nuclear myosin I (NMI). Cells were also treated with the marine toxin Latrunculin A (Lat-A), which affects the function of the nuclear actin. Under both conditions we observed a decrease in the percentage of compartments exhibiting directed motion as well as an increase in the percentage of compartments exhibiting obstructed diffusion. To further analyze the influence of nuclear myosin I (NMI) on the motion of the compartments, cells were transfected with the mutant NMI-E407V, which has an impaired motor activity. Under this condition again a decrease in the percentage of compartments exhibiting directed motion was observed. To further analyze the influence of nuclear actin, cells were transfected with the mutant actin-G13, which inhibits long-range movement of chromosomal loci. This condition also led to a decrease in the number of compartments undergoing directed motion. Cells were also transfected with a nuclear-targeted wild-type actin (actin-NLS), which also led to a reduction of the number of compartments exhibiting directed motion. The statistics obtained for each condition are shown in Figure 5.45. In addition, transcription was studied. Transcription is a process where a portion of a DNA nucleotide sequence (a gene) is copied into a RNA nucleotide sequence [3]. Transcription plays presumably a role in the movement of chromosomal loci. Cells were treated with the RNA pol II

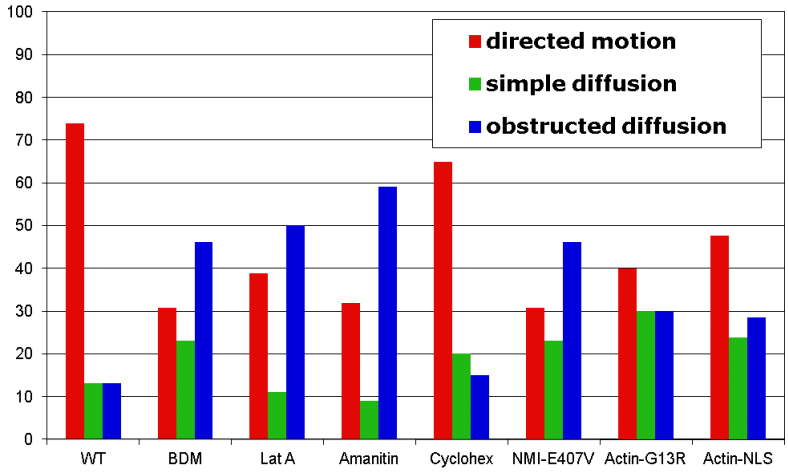


Figure 5.45: Distribution of motion types of HSV replication compartments under different experimental conditions. Trajectories were classified as directed motion (red), simple diffusion (green), or obstructed diffusion (blue). The distributions of motion types are displayed as a percentage of the total number of replication compartments in each condition.

inhibitor, α-amanitin, which also led to a decrease in the percentage of compartments exhibiting directed motion. In contrast, the protein synthesis inhibitor, cycloheximide, did not lead to a decrease in the number of compartments exhibiting directed motion. These results indicate that the motion of HSV replication compartments depends on nuclear actin, myosin, and ongoing transcription [35].

5.4 Tracking Microtubule Tips in *Xenopus laevis* Neurons

We have applied our tracking approaches to 2D image sequences displaying microtubule ends (tips) in nerve cells (neurons) [122]. Neural development drives the formation of the nervous system. Within the formation process, neurons connect to each other thereby forming a network of neurons. The neuron thus receives and transmits signals to other neurons. Within a single neuron, signals are transmitted over a large fiber called the *axon*. During the growth phase of the neuron, the axon extends and at the tip of the axon a cone-like structure is visible [3]. The dynamical behavior of the *growth cone* plays an important role in neural development. Microtubules (MT) are involved in the guidance of the growth cone. The understanding of the precise role of MT dynamics in axon growth is limited. Fluorescent labeling of the MT ends allow observing the motion of the MT ends during the growth process. Tracking the MT ends allow an objective quantification of the motion properties of the MT ends under different

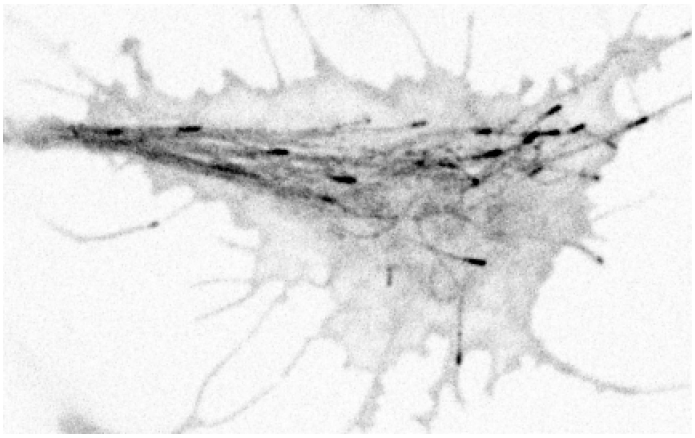


Figure 5.46: Real microscopy images displaying MT tips (dark spots) within a growth cone. Image intensities have been inverted for visualization purposes. The time step is $t = 100$.

conditions. Microtubules, being elongated cylindrical-like structures, have two ends and these are denoted as - and + ends, which denote their 'polarity'. In our case, the + ends were labeled with the end-binding protein 3 (EB3-mCherry). The focus is to study the relation of MT motion and the cytoplasmic linker associated protein (XCLASP1) in *Xenopus laevis* (African clawed frog) neurons. Images were acquired with a Nikon TiE microscope equipped with a Hamamatsu ORCA CCD camera at a rate of one image every 3 seconds. The dimensions of the images are typically 512×512 pixels (16-bit) but larger images are also acquired (e.g., 1344×1024 pixels). A sample image is shown in Figure 5.46. We have analyzed 122 image sequences in total.

To track the microtubuli tips we applied the PDAE approach (see Section 3.7). We used an anisotropic 2D Gaussian function to describe the appearance of a particle (cf. Section 3.8.2.1). We used a directed motion model (cf. Section 3.8.2.2) as well as circular sector measurements (cf. Section 3.8.2.4) to cope with the dynamic stability of the microtubules. 2D Gaussian fitting was used for localization. Sample tracking results are shown in Figure 5.47. It can be seen that the approach yields good results.

Based on the tracking results, parameters describing the motion of individual MT tips were computed, viz: mean displacement, MT lifetime (as reflected by the temporal duration of the corresponding trajectory), as well as the spatial extent covered during the motion of the MT. The statistics revealed the motion patterns of the MT tips in different parts of the neuron: axon, growth cone, as well as filamentous protrusions coming out of the growth cone (filopodia). It turned out that the microtubules slow as they transit into the growth cone's periphery. The motion statistics obtained under control and treatment conditions were compared. Cells were treated with the taxol drug and this slowed down the MT tips within the growth cone as well as in the axon shaft.

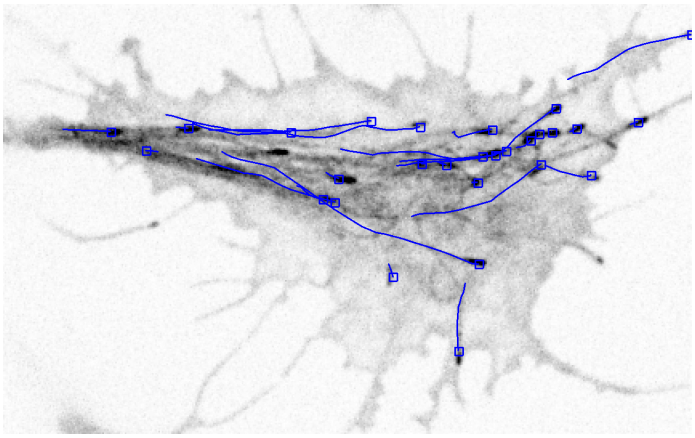


Figure 5.47: Tracking results for real microscopy images displaying MT tips. Trajectories are displayed in blue. Images have been inverted for visualization purposes. The time step is $t = 100$.

Thus taxol affects MT motion. In addition, the role of XCLASP1 was investigated via a loss-of-function approach. XCLASP1 was depleted in the cells by injecting a morpholino oligonucleotide. Under this condition, it turned out that the speed of the MTs is significantly reduced. Similar to the taxol treatment, the advance of the MTs into the periphery of the growth cone is inhibited and this correlates with a decreased axon motion. Thus it appears that XCLASP1 promotes the motion of MTs, in particular, the motion of those MTs going into the periphery of the growth cone [122].

5.5 Evaluation on the 2D and 3D Image Data of the ISBI'2012 Particle Tracking Challenge

We have also applied our PDAE tracking approach to the image data of the Particle Tracking Challenge held in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI) 2012 in Barcelona, Spain (http://www.bioimageanalysis.org/track/). The challenge comprises four different application scenarios: vesicles, virus particles, receptors, and microtubule tips. Realistic synthetic image sequences were generated for each scenario with different levels of image noise and object densities. For the virus particles, 3D image data was available and for the other scenarios 2D image data. In total, the data consists of 48 image sequences with 100 time steps each. 14 different measures were computed to quantify the performance of the tracking approaches (e.g., RMSE and Jaccard similarity index for the positions [38]). For the images showing virus particles and receptors we used both a random walk model as well as a directed motion model. For the vesicle images we used a random

walk model while for the the microtubule images we used a directed motion model. In the Particle Tracking Challenge, 14 groups world-wide participated. It turned out that our approach yielded the best overall performance (result announced at ISBI'2012, a journal publication on the Particle Tracking Challenge is in preparation).

5.6 Summary

This chapter presented the experimental evaluation of our approaches for tracking as well as for behavior identification. The evaluations were conducted based on synthetic images as well as real image sequences displaying HIV-1 particles. The performance of the approaches was quantified. Experimental comparisons with other approaches have also been carried out. In general, the experimental results show the good performance of our approaches. We also applied the approaches to 3D microscopy images displaying HSV replication compartments. From this study it turned out that certain nuclear mechanisms play a role in the motion of the compartments. Tracking results on images displaying microtubuli tips were also presented. In this study it was found that a certain protein plays an influential role in the motion of the microtubules within a neuron's growth cone.

Chapter 6

Conclusion

In this thesis, we considered the problem of estimating behaviors of fluorescent particles based on time-lapse microscopy image sequences. We decomposed this problem into three subtasks, viz. object localization, tracking, as well as behavior identification. We focused on the tasks of tracking, in particular on spatial-temporal filtering for position estimation, as well as on behavior identification. Below, we summarize our approaches, discuss our experimental results, and describe possible future work.

6.1 Summary

We developed novel probabilistic approaches grounded within a Bayesian framework for both tracking as well as behavior identification. In particular, we developed an efficient and robust approach based on probabilistic data association and an ellipsoidal sampling scheme (PDAE). We also developed a layered probabilistic approach for behavior identification based on the intensity over time of individual fluorescent particles.

More specifically, we have developed fully automatic probabilistic approaches for virus tracking in time-lapse fluorescence microscopy images. We have also performed an extensive evaluation of deterministic and probabilistic tracking approaches based on real microscopy image sequences displaying HIV-1 particles. The deterministic approaches follow a classic two-step paradigm, while the probabilistic approaches are based on Kalman filters and particle filters. Our extensive evaluation based on real image sequences indicate that the deterministic approaches have problems under realistic imaging situations (e.g., spurious particles). This arises mainly because of errors in both the localization algorithm (e.g., detection failures) and the motion correspondence step (e.g., incorrect assignments). The overall tracking accuracy can be enhanced by including a spatial-temporal filtering step. This is demonstrated via the superior results by the Kalman filter, and an even higher tracking accuracy is achieved with particle filters, in

particular, with independent particle filters. The Kalman filter relies on a deterministic particle localization approach, which diminishes its performance in cases where object localization is difficult. In contrast, the approaches based on particle filters are more robust and accurate under such conditions. The reason is mainly due to the comprehensive tracking machinery of the particle filter, which includes the steps of particle localization, motion correspondence, and position estimation. The particle filter manages to localize particles with poor contrast by directly analyzing the image data. In addition, the good performance of our approach based on independent particle filters in combination with our penalization scheme suggests that this approach is well-suited for multiple object tracking. Moreover, our multiple-channel measurement model enables handling of multi-channel microscopy image sequences.

A disadvantage of our approach based on independent particle filters is that the computational costs are relatively high and thus the approach is not well suited for biological applications with a large number of particles (e.g., 500 particles). For this reason, we have introduced a new approach for tracking multiple particles in fluorescence microscopy image sequences based on probabilistic data association. The new approach embraces the principles underlying the particle filter (e.g., multiple measurements, recognition-by-synthesis, use of parametric appearance models, direct use of the image data) in a more efficient manner. We have proposed a localization approach based on both a bottom-up strategy using a spot detection scheme (e.g., the spot-enhancing filter) as well as a top-down strategy using an elliptical sampling scheme for exploring the position space. The multiple measurements generated by our localization approach are integrated via the combined innovation principle of the probabilistic data association algorithm. For calculating the combined innovation, a weight is assigned to each measurement, and this weight is obtained by querying an image likelihood. In our case, the image likelihood considers both the predicted appearance of the object as well as the observed image intensities, and thus our approach uses the image information directly. For tracking multiple objects, we calculate the support that each image position provides to a tracked object relative to the support that the position provides to its neighboring objects. To incorporate multiple motion models, we used the interacting multiple model (IMM) algorithm. In comparison to tracking approaches based on the standard Kalman filter, our approach does not rely solely on a single measurement generated by a spot detection scheme and assigned by a motion correspondence algorithm. Instead, in addition to the measurement generated by a spot detection scheme, our approach steadily supplies measurements to the Kalman filter via the proposed top-down ellipsoidal sampling scheme. Thus, our approach is robust to errors arising from the spot detection scheme and in the correspondence algorithm. Moreover, our approach

considers the appearance of individual objects and uses the image information directly, i.e., segmentation of a particle is not necessarily required for estimating the position over time. For the approach based on independent particle filters, a relatively low number of samples (e.g., $N_s = 130$ in 2D) is not sufficient to obtain an accurate numerical approximation of the posterior probabilities. In comparison, the PDAE approach is not subject to such numerical issues since the posterior probabilities are represented analytically via the underlying Kalman filter, and so the PDAE approach operates well with a relatively low number of samples. Also, the ellipsoidal sampling scheme explores the position space more efficiently in comparison to the random strategy used by the particle filter. In addition to the parameters of the standard Kalman filter (e.g., the covariance matrices determining the noise processes) and of the IMM algorithm (e.g., transition probabilities), the PDAE approach is parametrized by the image noise parameter σ_n as well as by the discretization parameters N_c, N_j, and N_k that determine the number of measurements generated by the ellipsoidal sampling scheme. The noise parameter σ_n is adapted based on the measured image intensities for the particles. For the discretization parameters we have used fixed values for all our experiments.

We have applied the PDAE approach to synthetic images as well to real microscopy images and carried out an experimental comparison with approaches based on the standard Kalman filter as well as based on independent particle filters (IPF). Using synthetic images, we explored different levels of image noise, and characterized the localization performance of the approaches as a function of the SNR. For high SNR levels, the PDAE approach outperformed the other approaches. For low SNR levels, the PDAE approach performed better than the Kalman filter, and yielded a similar performance as the IPF approach. In addition, by varying the number of objects in synthetic image sequences, we characterized the performance of the approaches as a function of the probability of overlap between objects. It turned out that the performance of the approaches decreased linearly with the probability of overlap. Overall, the proposed PDAE approach outperformed the other approaches. We also determined the computation time of all approaches relative to the number of objects. The Kalman filter induced the lowest computation time, while the IPF entailed the highest computation time. The PDAE approach was somewhat slower than the Kalman filter (factor of 3) but significantly faster than the particle filter (factor of about 8). The application of the PDAE approach to real microscopy image data showed its suitability for tracking HIV-1 particles in 2D as well as 3D real microscopy images. Compared to the Kalman filter and the IPF, the PDAE approach yielded the best tracking accuracy on average.

We have also introduced a new approach for the identification of behaviors of interest for single virus particles in two-channel fluorescence microscopy image sequences. Our

approach is based on a layered architecture formulated within a Bayesian estimation framework. We use a stochastic hybrid system for modeling the intensity and the temporal intensity models. Our approach also employs a hidden Markov model (HMM) to represent the top-level behaviors of a single particle. The stochastic hybrid system and the HMM are combined using a maxbelief approach. Inference is carried out with a hybrid particle filter as well as with the Viterbi algorithm. In comparison to previous approaches for identifying behaviors of fluorescent particles, the approach does not assume a constant behavior for the particles, and thus the approach can cope with objects exhibiting heterogeneous behaviors. Also, we take into account the temporal coherence of the particle's behavior via the transition probabilities defined for the Markov chains underlying the stochastic hybrid system and the HMM. By adopting a probabilistic framework, we account for the inherent uncertainty involved in the behavior of single virus particles.

We have applied our layered approach to synthetic image sequences as well as to real image sequences. We also have compared the performance of our approach with that of a previous derivative-based approach. In addition, we have performed a comparison with an approach based on trajectory endpoints. The results on two-channel synthetic images demonstrated the applicability of our layered approach for identifying the intensity models as well as the behavior of single particles. By exploring a range of different levels of noise, we characterized the performance of our approach as a function of the SNR level. The results established that our approach performed well at typical SNR levels and relatively well at low SNR levels. Also the performance of the approach degraded gracefully as the SNR decreased. In addition, we evaluated the performance of the approach relative to the performance of the tracking approach. It turned out that the layered approach tolerates well minor tracking errors and that the performance degrades smoothly with larger tracking errors. The application of the layered approach to real microscopy image data showed that the approach is well suited for detecting fusion of HIV-1 particles with the cell membrane. Overall, the layered approach yielded a significant improvement in the performance compared to a derivative-based approach, which relies entirely on data-driven information. Instead, our approach integrates model-driven information (e.g., predictions based on the temporal intensity models) with the observed intensity information. The layered approach also outperformed an approach using trajectory endpoints. This highlights the benefits of directly analyzing the image intensities of individual particles. Certainly, the layered approach has also limitations. Since a compromise between the adaptation time and the steady behavior of the hybrid particle filter has to be found, the labels computed by the approach may exhibit a certain time delay. This time delay can be reduced by adjusting the transition prob-

abilities in the matrix $\mathbf{\Pi}$, the variance parameters Q_α as well as the noise parameter σ_n. However, doing this for single image sequences might be impractical. Alternatively, these parameters could be optimized based on the image data.

Besides considering HIV-1 particles, we have additionally used our tracking approaches in other biological applications involving herpes simplex virus (HSV) replication compartments as well as microtubuli tips within the neural growth cone. For analyzing the motion of the HSV compartments, we developed a hierarchical approach. Through the results obtained with our approaches, novel biological findings were obtained. The results also demonstrate the capability of the approaches for handling different types of biological structures.

We also successfully applied our PDAE approach to the 2D and 3D image data of the recent Particle Tracking Challenge at the International Symposium on Biomedical Imaging (ISBI) 2012, where 14 groups world-wide participated. The challenge comprised four different application scenarios (vesicles, virus particles, receptors, and microtubule tips) and different performance measures were computed. It turned out, that the PDAE approach yielded the best overall performance (result announced at ISBI'2012).

6.2 Future Work

The PDAE tracking approach currently uses the temporal information sequentially. One extension would be using the temporal information from the entire image sequence to improve the performance. Another issue is related to the high level of cellular autofluorescence, which leads to spurious particles, thereby diminishing the performance of the bottom-up detection scheme. Here, a more accurate recognition mechanism could be used. The performance for larger object overlaps could also be improved by using a more sophisticated algorithm for motion correspondence. Using image data of multiple microscopy modalities (e.g., fluorescence microscopy, bright field microscopy) could also improve the performance, since the image data of each modality might complement each other. Instead of using the Kalman filter, one could also use one of its non-linear variants (e.g., the unscented Kalman filter) which would allow incorporating non-linear dynamical models. Finally, the PDAE approach could be extended to cope with photobleaching effects by estimating the rate of change of the intensity over time.

Concerning the layered the approach for behavior identification, one issue is to improve the performance of the approach in case of severe tracking errors. Here the layered approach would also benefit from using the entire temporal information as opposed to only using the temporal information sequentially. Also, the set of models describing the temporal statistics as well as the behavior could be adjusted over time. Developing

automatic schemes for adjusting the set of models as well as for automatically designing new models would be another research avenue. The layered approach could also be applied in other applications using image sequences with more than two channels. Current and future work includes the further interpretation of the results delivered by the layered approach in terms of their biological significance.

Bibliography

[1] M. Abramoff, P. Magelhaes, and S. Ram, "Image Processing with ImageJ", *Biophotonics International*, vol. 11, no. 7, pp. 36–42, 2004.

[2] D. Agard, Y. Hiraoka, P. Shaw, and J. Sedat, "Fluorescence Microscopy in Three Dimensions", in *Fluorescence Microscopy of Living Cells in Culture Part B. Quantitative Fluorescence Microscopy–Imaging and Spectroscopy*, vol. 30, Elsevier, 1989, pp. 353–377.

[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*, 4th ed. New York, New York: Garland Science Taylor & Francis Group, 2002.

[4] A. Ali and K. Terada, "A framework for human tracking using Kalman filter and fast mean shift algorithms", in *Proc. IEEE Int. Conf. on Computer Vision Workshops (ICCVW'09)*, Kyoto, Japan, Oct. 2009, pp. 1028–1033.

[5] B. Anderson and J. Moore, *Optimal Filtering*. Mineola, NY, USA: Dover Publications, 1979, pp. 9–34.

[6] C. Anderson, G. Georgiou, I. Morrison, G. Stevenson, and R. Cherry, "Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera. Low-density lipoprotein and influenza virus receptor mobility at 4 degrees C", *J. Cell Sci.*, vol. 101, no. 2, pp. 415–425, 1992.

[7] H. Andersson, T. Baechi, M. Hoechl, and C. Richter, "Autofluorescence of living cells", *J. Microsc.*, vol. 191, no. 1, pp. 1–7, 1998.

[8] C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions", *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1762–1770, 2003.

[9] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking", *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.

[10] C. Bacher, M. Reichenzeller, C. Athale, H. Herrmann, and R. Eils, "4-D single particle tracking of synthetic and proteinaceous microspheres reveals preferential movement of nuclear particles along chromatin - poor tracks", *BMC Cell Biol.*, vol. 5, no. 1, p. 45, 2004.

[11] L. Bai, Y. Wang, J. Fan, Y. Chen, W. Ji, A. Qu, P. Xu, D. E. James, and T. Xu, "Dissecting multiple steps of GLUT4 trafficking and identifying the sites of insulin action", *Cell Metab.*, vol. 5, no. 1, pp. 47–57, 2007.

[12] H.-J. Bandelt and A. Roehl, "Quasi-median hulls in Hamming space are Steiner hulls", *Discrete Appl. Math.*, vol. 157, no. 2, pp. 227–233, 2009.

[13] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Object trajectory-based activity classification and recognition using hidden markov models", *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1912–1919, 2007.

[14] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using K-shortest paths optimization", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1806–1819, 2011.

[15] S. M. Bhandarkar and X. Luo, "Integrated detection and tracking of multiple faces using particle filtering and optical flow-based elastic matching", *Comput. Vision Image Understanding*, vol. 113, no. 6, pp. 708–725, 2009.

[16] N. Billinton and A. W. Knight, "Seeing the Wood through the Trees: A Review of Techniques for Distinguishing Green Fluorescent Protein from Endogenous Autofluorescence", *Anal. Biochem.*, vol. 291, no. 2, pp. 175–197, 2001.

[17] A. Bissacco and S. Soatto, "Hybrid dynamical models of human motion for the recognition of human gaits", *Int. J. Comput. Vision*, vol. 85, pp. 101–114, 1 2009.

[18] M. J. Black and A. D. Jepson, "A probabilistic framework for matching temporal trajectories: CONDENSATION-based recognition of gestures and expressions", in *Proc. European Conf. on Computer Vision (ECCV'98)*, Freiburg, Germany, Jun. 1998, pp. 909–924.

[19] S. Blackman, "Multiple hypothesis tracking for multiple target tracking", *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, 2004.

[20] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'05)*, Beijing, China, Oct. 2005, pp. 1395–1402.

[21] H. A. P. Blom, "An efficient filter for abruptly changing systems", in *Proc. IEEE Conf. on Decision and Control*, Las Vegas, NV, USA, Dec. 1984, pp. 656–658.

[22] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients", *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, 1988.

[23] S. Bonneau, M. Dahan, and L. D. Cohen, "Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume", *IEEE Trans. Image Process.*, vol. 14, pp. 1384–1395, 2005.

[24] J. Boulanger, A. Gidon, C. Kervran, and J. Salamero, "A patch-based method for repetitive and transient event detection in fluorescence imaging", *PLoS ONE*, vol. 5, no. 10, e13190, 2010.

[25] E. Brau, K. Barnard, R. Palanivelu, D. Dunatunga, T. Tsukamoto, and P. Lee, "A generative statistical model for tracking multiple smooth trajectories", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, Colorado Springs, CO, USA, Jun. 2011, pp. 1137–1144.

[26] C. Bregler, "Learning and recognizing human dynamics in video sequences", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, Jun. 1997, pp. 568–574.

[27] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, pp. 1820–1833, 2011.

[28] D. S. Bright and E. B. Steel, "Two-dimensional top hat filter for extracting spots and spheres from digital images", *J. Microsc.*, vol. 146, no. 2, pp. 191–200, 1987.

[29] S. Butail and D. Paley, "3D reconstruction of fish schooling kinematics from underwater video", in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA'10)*, Anchorage, AK, USA, May 2010, pp. 2438–2443.

[30] Y. Cai, N. de Freitas, and J. J. Little, "Robust visual tracking for multiple targets", in *Proc. European Conf. on Computer Vision (ECCV'06)*, Graz, Austria, May 2006, pp. 107–118.

[31] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding transit scenes: a survey on human behavior-recognition algorithms", *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 206–224, 2010.

[32] J. Cardinale, A. Rauch, Y. Barral, G. Székely, and I. F. Sbalzarini, "Bayesian image analysis with on-line confidence estimates and its application to microtubule tracking", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston, MA, USA, Jun. 2009, pp. 1091–1094.

[33] C. Cédras and M. A. Shah, "Motion based recognition: A survey.", *Image Vision Comput.*, vol. 13, no. 2, pp. 129–155, Mar. 1995.

[34] T.-J. Cham and J. M. Rehg, "A multiple hypothesis approach to figure tracking.", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'99)*, Fort Collins, Colorado, USA, Jun. 1999, pp. 2239–2245.

[35] L. Chang, W. J. Godinez, I.-H. Kim, M. Tektonidis, P. de Lanerolle, R. Eils, K. Rohr, and D. M. Knipe, "Herpesviral replication compartments move and coalesce at nuclear speckles to enhance export of viral late mRNA", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 21, E136–E144, 2011.

[36] M. K. Cheezum, W. F. Walker, and W. H. Guilford, "Quantitative comparison of algorithms for tracking single fluorescent particles", *Biophys. J.*, vol. 81, pp. 2378–2388, 2001.

[37] J. Chen, M. Kim, Y. Wang, and Q. Ji, "Switching Gaussian process dynamic models for simultaneous composite motion tracking and recognition", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'09)*, Miami, FL, USA, Jun. 2009, pp. 2655–2662.

[38] N. Chenouard, I. Bloch, and J.-C. Olivo-Marin, "Multiple hypothesis tracking in microscopy images", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston, MA, USA, Jun. 2009, pp. 1346–1349.

[39] L. S. Churchman, Z. Okten, R. S. Rock, J. F. Dawson, and J. A. Spudich, "Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 5, pp. 1419–1423, 2005.

[40] C. Collinet, M. Stoter, C. R. Bradshaw, N. Samusik, J. C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M. S. Mueller, W. E. Nagel, E. Fava, Y. Kalaidzidis, and M. Zerial, "Systems survey of endocytosis by multiparametric image analysis", *Nature*, vol. 464, no. 7286, pp. 243–249, 2010.

[41] L. Cortes and Y. Amit, "Efficient annotation of vesicle dynamics in video microscopy", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1998–2010, 2008.

[42] I. J. Cox, "A review of statistical data association for motion correspondence", *Int. J. Comput. Vision*, vol. 10, no. 1, pp. 53–66, 1993.

[43] I. J. Cox and S. L. Hingorani, "An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, 1996.

[44] J. Cui, S. T. Acton, and Z. Lin, "A Monte Carlo approach to rolling leukocyte tracking in vivo", *Med. Image Anal.*, vol. 10, no. 4, pp. 598–610, 2006.

[45] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[46] G. Dalley, X. Wang, and E. Grimson, "Event detection using an attention-based tracker", in *Proc. IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (VS-PETS'05)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 71–78.

[47] S. B. Dalziel, "Decay of rotating turbulence: some particle tracking experiments", *App. Sci. Res.*, vol. 49, no. 3, pp. 217–244, 1992.

[48] J. Darby, B. Li, and N. Costen, "Tracking human pose with multiple activity models", *Pattern Recognit.*, vol. 43, no. 9, pp. 3042–3058, 2010.

[49] M. Das and R. Manmatha, "Automatic segmentation and indexing in a database of bird images", *IEEE Int. Conf. on Computer Vision (ICCV'01)*, pp. 351–358, Jul. 2001.

[50] T. De Laet, H. Bruyninckx, and J. De Schutter, "Shape-based online multitarget tracking and detection for targets causing multiple measurements: variational bayesian clustering and lossless data association", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2477–2491, 2011.

[51] N. Deng, Y. Xu, D. Sun, P. Hua, X. Zheng, and H. Duan, "Image processing for fusion identification between the GLUT4 storage vesicles and the plasma membrane", *J. Signal Process. Syst.*, vol. 54, pp. 115–125, 1-3 2009.

[52] A. Djouguela, D. Harder, R. Kollhoff, S. Foschepoth, W. Kunth, A. Ruehmann, K. Willborn, and B. Poppe, "Fourier deconvolution reveals the role of the Lorentz function as the convolution kernel of narrow photon beams", *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2807–2827, 2009.

[53] P. Dollar, V. Rabaud, G. Cottrel, and S. Belongie, "Behavior recognition via spatio-temporal features", in *Proc. IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS'05)*, Beijing, China, Oct. 2005, pp. 65–72.

[54] F. Dornaika and F. Davoine, "Simultaneous facial action tracking and expression recognition in the presence of head motion", *Int. J. Comput. Vision*, vol. 76, no. 3, pp. 257–281, 2008.

[55] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to Sequential Monte Carlo methods", in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., New York, Berlin, Heidelberg: Springer-Verlag, 2001, pp. 3–14.

[56] A. Doucet, "On Sequential Simulation-Based Methods for Bayesian Filtering", Cambridge University Department of Engineering, Tech. Rep. CUED/F-INFENG/TR. 310, 1998.

[57] H. Driessen and Y. Boers, "An efficient particle filter for jump Markov nonlinear systems", in *Proc. IEE Target Tracking 2004: Algorithms and Applications*, Brighton, UK, Mar. 2004, pp. 19–22.

[58] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'03)*, Nice, France, Oct. 2003, pp. 726–733.

[59] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: survey and experiments", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, 2009.

[60] L. Feng, Y. Xu, Y. Yang, and X. Zheng, "Multiple dense particle tracking in fluorescence microscopy images based on multidimensional assignment", *J. Struct. Biol.*, vol. 173, no. 2, pp. 219–228, 2011.

[61] D. L. Floyd, J. R. Ragains, J. J. Skehel, S. C. Harrison, and A. M. van Oijen, "Single-particle kinetics of influenza virus membrane fusion", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, no. 40, pp. 15 382–15 387, 2008.

[62] G. Forney, "The Viterbi algorithm", *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[63] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle", *Int. J. Comput. Vision*, vol. 73, no. 1, pp. 41–59, 2007.

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

[65] A. Genovesio and J.-C. Olivo-Marin, "Split and merge data association filter for dense multi-target tracking", in *Proc. Int. Conf. on Pattern Recognition (ICPR'04)*, vol. 4, Cambridge, UK, Aug. 2004, pp. 677–680.

[66] D. Gerónimo, A. López, A. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, 2010.

[67] J. Giebel, D. Gavrila, and C. Schnörr, "A Bayesian framework for multi-cue 3D object tracking", in *Proc. European Conf. on Computer Vision (ECCV'04)*, Prague, Czech Republic, May 2004, pp. 241–252.

[68] W. J. Godinez, M. Lampe, R. Eils, B. Müller, and K. Rohr, "Tracking multiple particles in fluorescence microscopy images via probabilistic data association", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'11)*, Chicago, IL, USA, Mar. 2011, pp. 1925–1928.

[69] W. J. Godinez, M. Lampe, P. Koch, R. Eils, B. Müller, and K. Rohr, "Identifying virus-cell fusion in two-channel fluorescence microscopy image sequences based on a layered probabilistic approach", *IEEE Trans. Med. Imaging*, vol. 31, no. 9, pp. 1786–1808, 2012.

[70] W. J. Godinez, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, "Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences", *Med. Image Anal.*, vol. 13, no. 2, pp. 325–342, 2009.

[71] W. J. Godinez, "Tracking of Virus Particles in Time-Lapse Fluorescence Microscopy Image Sequences", Master's thesis, International University in Germany, Bruchsal, Germany, 2007.

[72] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proc. F*, vol. 140, no. 2, pp. 107–113, 1993.

[73] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[74] O. Gress and S. Posch, "Trajectory retrieval from Monte Carlo data association samples for tracking in fluorescence microscopy images", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, May 2012, pp. 374–377.

[75] H. Gribben, P. Miller, J. Zhang, and M. Browne, "Poisson Kalman particle filtering for tracking centrosomes in low-light 3-D confocal image sequences", in *Proc. Int. Machine Vision and Image Processing Conf. (IMVIP'09)*, Dublin, Ireland, Sep. 2009, pp. 83–88.

[76] L. Gui, J.-P. Thiran, and N. Paragios, "Cooperative object segmentation and behavior inference in image sequences", *Int. Journal of Computer Vision*, vol. 84, no. 2, pp. 146–162, 2009.

[77] D. W. Hansen and Q. Ji, "In the eye of the beholder: a survey of models for eyes and gaze", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, 2010.

[78] R. Haralick and L. Shapiro, *Computer and Robot Vision*. Boston, MA, USA: Addison-Wesley, 1992, vol. 1.

[79] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[80] S. Hell, "Improvement of lateral resolution in far-field fluorescence light microscopy by using two-photon excitation with offset beams", *Opt. Comm.*, vol. 106, no. 1-3, pp. 19–24, 1994.

[81] G. A. Herman, F. Bonzelius, A. M. Cieutat, and R. B. Kelly, "A Distinct Class of Intracellular Storage Vesicles, Identified by Expression of the Glucose Transporter GLUT4", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 26, 1994.

[82] M. Hofbaur and B. Williams, "Hybrid estimation of complex systems", *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 34, no. 5, pp. 2178–2191, 2004.

[83] T. M. Hospedales, J. Li, S. Gong, and T. Xiang, "Identifying rare and subtle behaviors: a weakly supervised joint topic model", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2451–2464, 2011.

[84] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors", *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 34, no. 3, pp. 334–352, 2004.

[85] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. J. Maybank, "A system for learning statistical motion patterns", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, 2006.

[86] C. Huang, B. Wu, and R. Nevatia, "Robust object tracking by hierarchical association of detection responses", in *Proc. European Conf. on Computer Vision (ECCV'08)*, Marseille, France, Oct. 2008, pp. 788–801.

[87] M. Isard and A. Blake, "A mixed-state CONDENSATION tracker with automatic model-switching", in *Proc. Int. Conf. on Computer Vision (ICCV'98)*, Bombay, India, Jan. 1998, pp. 107–112.

[88] ——, "CONDENSATION – Conditional density propagation for visual tracking", *Int. J. Comput. Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[89] M. Isard and J. MacCormick, "BraMBLe: a Bayesian multiple-blob tracker.", in *Proc. Int. Conf. on Computer Vision (ICCV'01)*, Vancouver, BC, Canada, Jul. 2001, pp. 34–41.

[90] S. Ivanchenko, W. J. Godinez, M. Lampe, H.-G. Kräusslich, R. Eils, K. Rohr, C. Bräuchle, B. Müller, and D. C. Lamb, "Dynamics of HIV-1 assembly and release", *PLoS Pathog.*, vol. 5, no. 11, e1000652, Nov. 2009.

[91] S. Iwase and H. Saito, "Parallel tracking of all soccer players by integrating detected positions in multiple view images", in *Proc. Int. Conf. on Pattern Recognition (ICPR'04)*, Cambridge, UK, Aug. 2004, pp. 751–754.

[92] S. Jaensch, M. Decker, A. A. Hyman, and E. W. Myers, "Automated tracking and analysis of centrosomes in early Caenorhabditis elegans embryos", *Bioinformatics*, vol. 26, no. 12, pp. i13–i20, 2010.

[93] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, "Robust single-particle tracking in live-cell time-lapse sequences", *Nat. Methods*, vol. 5, no. 8, pp. 695–702, 2008.

[94] H. Jiang, S. Fels, and J. J. Little, "A linear programming approach for multiple object tracking", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'07)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[95] S. Jiang, X. Zhou, T. Kirchhausen, and S. T. C. Wong, "Detection of molecular particles in live cells via machine learning", *Cytometry A*, vol. 71A, no. 8, pp. 563–575, 2007.

[96] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation", *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[97] R. E. Kalman, "A new approach to linear filtering and prediction problems", *J. Basic Eng.-T. ASME*, pp. 35–45, 1960.

[98] R. A. Kerekes, S. S. Gleason, N. Trivedi, and D. Solecki, "Automated 3-D tracking of centrosomes in sequences of confocal image stacks", in *Proc. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC'09)*, Minneapolis, MN, USA, Sep. 2009, pp. 6994–6997.

[99] Z. Khan, T. Balch, and F. Dellaert, "A Rao-Blackwellized particle filter for eigentracking", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 2, Washington, DC, USA, Jun. 2004, pp. 980–986.

[100] ——, "MCMC-based particle filtering for tracking a variable number of interacting targets", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805–1918, 2005.

[101] R. Kidambi, M.-C. Shih, and K. Rose, "Deformable trellises on factor graphs for robust microtubule tracking in clutter", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, May 2012, pp. 676–679.

[102] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter", *Proc. IEEE*, vol. 92, no. 3, pp. 536–557, 2004.

[103] P. Koch, M. Lampe, W. J. Godinez, B. Müller, K. Rohr, H.-G. Kräusslich, and M. Lehmann, "Visualizing fusion of pseudotyped HIV-1 particles in real time by live cell microscopy", *Retrovirology*, vol. 6, no. 1, p. 84, 2009.

[104] T. Kues, R. Peters, and U. Kubitscheck, "Visualization and tracking of single protein molecules in the cell nucleus", *Biophys. J.*, vol. 80, no. 6, pp. 2954–2967, 2001.

[105] P. Kukura, H. Ewers, C. Muller, A. Renn, A. Helenius, and V. Sandoghdar, "High-speed nanoscopic tracking of the position and orientation of a single virus", *Nat. Methods*, vol. 6, no. 12, pp. 923–927, 2009.

[106] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, and Y. Sato, "Pose-invariant facial expression recognition using variable-intensity templates", *Int. J. Comput. Vision*, vol. 83, no. 2, pp. 178–194, 2009.

[107] M. Lakadamyali, M. J. Rust, H. P. Babcock, and X. Zhuang, "Visualizing infection of individual influenza viruses", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 16, pp. 9280–9285, 2003.

[108] M. Lampe, J. A. G. Briggs, T. Endress, B. Glass, S. Riegelsberger, H.-G. Kräusslich, D. C. Lamb, C. Bräuchle, and B. Müller, "Double-labelled HIV-1 particles for study of virus-cell interaction", *Virology*, vol. 360, no. 1, pp. 92–104, 2007.

[109] O. Lanz, "Approximate Bayesian multibody tracking", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1436–1449, 2006.

[110] I. Laptev, "On space-time interest points", *Int. J. Comput. Vision*, vol. 64, no. 2-3, pp. 107–123, 2005.

[111] I. Laptev and P. Pérez, "Retrieving actions in movies", in *Proc. Int. Conf. on Computer Vision (ICCV'07)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.

[112] G. M. Lauer and B. D. Walker, "Hepatitis C virus infection", *N. Engl. J. Med.*, vol. 345, no. 1, pp. 41–52, 2001.

[113] K. Li, E. D. Miller, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Cell population tracking and lineage construction with spatiotemporal context", *Med. Image Anal.*, vol. 12, no. 5, pp. 546–566, 2008.

[114] R. Li, T.-P. Tian, S. Sclaroff, and M.-H. Yang, "3D human motion tracking with a coordinated mixture of factor analyzers", *Int. J. Comput. Vision*, vol. 87, no. 1-2, pp. 170–190, 2010.

[115] L. Liang, H. Shen, P. D. Camilli, and J. S. Duncan, "Tracking clathrin coated pits with a multiple hypothesis based method", in *Proc. Conf. on Medical Image Computing Computer-Assisted Intervention (MICCAI'10)*, vol. 2, Beijing, China, Sep. 2010, pp. 315–322.

[116] L. Liang, H. Shen, P. D. Camilli, D. Toomre, and J. S. Duncan, "An expectation maximization based method for subcellular particle tracking using multi-angle tirf microscopy", in *Proc. Conf. on Medical Image Computing Computer-Assisted Intervention (MICCAI'11)*, vol. 2, Toronto, Canada, Sep. 2011, pp. 629–636.

[117] J. W. Lichtman and J. A. Conchello, "Fluorescence microscopy", *Nat. Methods*, vol. 2, no. 12, pp. 910–919, 2005.

[118] C. Lu and N. J. Ferrier, "Repetitive motion analysis: segmentation and event classification", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 258–263, 2004.

[119] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'81)*, Vancouver, BC, Canada, Aug. 1981, pp. 674–679.

[120] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects", *Int. J. Comput. Vision*, vol. 39, no. 1, pp. 57–71, 2000.

[121] M. Marsh and A. Helenius, "Virus entry: open sesame", *Cell*, vol. 124, no. 4, pp. 729–740, 2006.

[122] A. Marx, W. J. Godinez, V. Tsimashchuk, P. Bankhead, K. Rohr, and U. Engel, "*Xenopus* cytoplasmic linker associated protein 1 (XCLASP1) promotes axon elongation and advance of pioneer microtubules", *Mol. Biol. Cell*, 2013, in press.

[123] A. Matov, K. Applegate, P. Kumar, C. Thoma, W. Krek, G. Danuser, and T. Wittmann, "Analysis of microtubule dynamic instability using a plus-end growth marker", *Nat. Methods*, vol. 7, no. 9, pp. 761–768, 2010.

[124] S. McGinnity and G. Irwin, "Multiple model bootstrap filter for maneuvering target tracking", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 3, pp. 1006–1012, 2000.

[125] K. Mele, A. Coster, J. Burchfield, J. Lopez, D. James, W. Hughes, and P. Vallotton, "Automatic identification of fusion events in TIRF microscopy image sequences", in *Proc. Int. Conf. on Computer Vision Workshops (ICCVW'09)*, Kyoto, Japan, Oct. 2009, pp. 578–584.

[126] K. Miyauchi, Y. Kim, O. Latinovic, V. Morozov, and G. B. Melikyan, "HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes", *Cell*, vol. 137, no. 3, pp. 433–444, 2009.

[127] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis", *Comput. Vision Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[128] B. Müller, J. Daecke, O. T. Fackler, M. T. Dittmar, H. Zentgraf, and H.-G. Kräusslich, "Construction and characterization of a fluorescently labeled infectious human immunodeficiency virus type 1 derivative", *J. Virol.*, vol. 78, no. 19, pp. 10 803–10 813, 2004.

[129] K. Murphy and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks", in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., New York, Berlin, Heidelberg: Springer-Verlag, 2001.

[130] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Trajectory classification using switched dynamical hidden markov models", *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1338–1348, 2010.

[131] N. T. Nguyen, H. H. Bui, S. Venkatesh, and G. A. W. West, "Recognising and monitoring high-level behaviours in complex spatial environments", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'09)*, Madison, WI, USA, Jun. 2003, pp. 620–625.

[132] W. Niles, Q. Li, and F. Cohen, "Computer detection of the rapid diffusion of fluorescent membrane fusion markers in images observed with video microscopy", *Biophys. J.*, vol. 63, no. 3, pp. 710–722, 1992.

[133] B. North, A. Blake, M. Isard, and J. Rittscher, "Learning and classification of complex dynamics", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 1016–1034, 2000.

[134] S. M. Oh, J. M. Rehg, T. R. Balch, and F. Dellaert, "Learning and inferring motion patterns using parametric segmental switching linear dynamic systems", *Int. J. Comput. Vision*, vol. 77, no. 1-3, pp. 103–124, 2008.

[135] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems", in *Proc. IEEE Conf. on Decision and Control (CDC'04)*, Nassau, Bahamas, Dec. 2004, pp. 735–742.

[136] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe, "A boosted particle filter: multitarget detection and tracking", in *Proc. European Conf. on Computer Vision (ECCV'04)*, Prague, Czech Republic, May 2004, pp. 28–39.

[137] N. Oliver, A. Garg, and E. Horvitz, "Layered representations for learning and inferring office activity from multiple sensory channels", *Comput. Vision Image Understanding*, vol. 96, no. 2, pp. 163–180, 2004.

[138] J.-C. Olivo-Marin, "Extraction of spots in biological images using multiscale products.", *Pattern Recognit.*, vol. 35, no. 9, pp. 1989–1996, 2002.

[139] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[140] K. P., A. Kokaram, J. Hillebrand, and M. Ramaswami, "Gaussian mixture models for spots in microscopy using a new split/merge EM algorithm", in *Proc. Int. Conf. on Image Processing (ICIP'10)*, Hong Kong, Sep. 2010, pp. 3645–3648.

[141] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization : Algorithms and Complexity*. Mineola, NY, USA: Dover Publications, 1998.

[142] A. Perera, C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu, "Multi-object tracking through simultaneous long occlusions and split-merge conditions", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, Jun. 2006, pp. 666–673.

[143] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, Colorado Springs, CO, USA, Jun. 2011, pp. 1201–1208.

[144] A. Ponti, P. Vallotton, W. C. Salmon, C. M. Waterman-Storer, and G. Danuser, "Computational analysis of f-actin turnover in cortical actin meshworks using fluorescent speckle microscopy", *Biophys. J.*, vol. 84, pp. 336–3352, 2003.

[145] R. Poppe, "A survey on vision-based human action recognition", *Image Vision Comput.*, vol. 28, no. 6, pp. 976–990, 2010.

[146] W. Qu, D. Schonfeld, and M. Mohamed, "Real-time interactively distributed multi-object tracking using a magnetic-inertia potential model", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'05)*, Beijing, China, Oct. 2005, pp. 535–540.

[147] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[148] C. Rao and M. Shah, "View-invariance in action recognition", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'10)*, Kauai, HI, USA, Dec. 2001, pp. 316–322.

[149] C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects.", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 23, no. 6, pp. 560–576, 2001.

[150] C. Rasmussen, "Gaussian processes in machine learning", in *Advanced Lectures on Machine Learning*, O. Bousquet, U. von Luxburg, and G. Raetsch, Eds., vol. 3176, New York, Berlin, Heidelberg: Springer-Verlag, 2004, pp. 63–71.

[151] Y. Rathi, N. Vaswani, and A. Tannenbaum, "A generic framework for tracking using particle filter with dynamic shape prior", *IEEE Trans. Image. Process.*, vol. 16, no. 5, pp. 1370–1382, 2007.

[152] D. B. Reid, "An Algorithm for Tracking Multiple Targets", *IEEE Trans. on Autom. Control*, no. 6, pp. 843–854, 1976.

[153] C. Restif, C. Ibañez-Ventoso, M. Driscoll, and D. N. Metaxas, "Tracking C. elegans swimming for high-throughput phenotyping", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'11)*, Chicago, IL, USA, Mar. 2011, pp. 1542–1548.

[154] S. H. Rezatofighi, S. Gould, R. Hartley, K. Mele, and W. E. Hughes, "Application of the IMM-JPDA filter to multiple target tracking in total internal reflection fluorescence microscopy images", in *Proc. Conf. on Medical Image Computing Computer-Assisted Intervention (MICCAI'12)*, Nice, France, Oct. 2012, pp. 357–364.

[155] J. Rink, E. Ghigo, Y. Kalaidzidis, and M. Zerial, "Rab conversion as a mechanism of progression from early to late endosomes", *Cell*, vol. 122, no. 5, pp. 735–749, 2005.

[156] J. Rittscher and A. Blake, "Classification of human body motion", in *Proc. Int. Conf. on Computer Vision (ICCV'99)*, Corfu, Greece, Sep. 1999, pp. 634–639.

[157] K. Rohr, "Towards model-based recognition of human movements in image sequences", *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 94–115, 1994.

[158] P. Ruusuvuori, T. Äijö, S. Chowdhury, C. Garmendia-Torres, J. Selinummi, M. Birbaumer, A. M. Dudley, L. Pelkmans, and O. Yli-Harja, "Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images", *BMC Bioinf.*, vol. 11, p. 248, 2010.

[159] S. S. Stalder, H. Grabner, and L. Van Gool, "Cascaded confidence filtering for improved tracking-by-detection", in *Proc. European Conf. on Computer Vision (ECCV'10)*, Heraklion, Crete, Greece, Sep. 2010, pp. 369–382.

[160] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser, "Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics", *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1372–1382, 2005.

[161] I. Saleemi, L. Hartung, and M. Shah, "Scene understanding by statistical modeling of motion patterns", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'10)*, San Francisco, CA, USA, Jun. 2010, pp. 2069–2076.

[162] M. Sargin, A. Altinok, B. Manjunath, and K. Rose, "Variable length open contour tracking using a deformable trellis", *IEEE Trans. Image. Process.*, vol. 20, no. 4, pp. 1023–1035, 2011.

[163] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-Blackwellized Monte Carlo data association for multiple target tracking", in *Proc. Int. Conf. on Information Fusion (FUSION'2004)*, Stockholm, Sweden, Jun. 2004, pp. 583–590.

[164] M. Saxton, "Lateral diffusion in an archipelago. Single-particle diffusion", *Biophys. J.*, vol. 64, no. 6, pp. 1766–1780, 1993.

[165] M. Saxton and K. Jacobson, "Single-particle tracking: Applications to membrane dynamics", *Annu. Rev. Biophys. Biomol. Struct.*, vol. 26, no. 1, pp. 373–399, 1997.

[166] I. F. Sbalzarini and P. Koumoutsakos, "Feature point tracking and trajectory analysis for video imaging in cell biology", *J. Struct. Biol.*, vol. 151, no. 2, pp. 182–195, 2005.

[167] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters", *Int. J. Rob. Res.*, vol. 22, no. 2, pp. 99–116, 2003.

[168] D. Schulz, D. Fox, and J. Hightower, "People tracking with anonymous and ID-sensors using Rao-Blackwellised particle filters", in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, Aug. 2003, pp. 921–926.

[169] G. J. Schutz, H. Schindler, and T. Schmidt, "Single-molecule microscopy on model membranes reveals anomalous diffusion", *Biophys. J.*, vol. 73, no. 2, pp. 1073–1080, 1997.

[170] R. Sebastian, M.-E. Diaz, G. Ayala, K. Letinic, J. Moncho-Bogani, and D. Toomre, "Spatio-temporal analysis of constitutive exocytosis in epithelial cells", *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 3, no. 1, pp. 17–32, 2006.

[171] G. Seisenberger, M. U. Ried, T. Endress, H. Buning, M. Hallek, and C. Bräuchle, "Real-time single-molecule imaging of the infection pathway of an adeno-associated virus", *Science*, vol. 294, no. 5548, pp. 1929–1932, 2001.

[172] A. Serge, N. Bertaux, H. Rigneault, and D. Marguet, "Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes", *Nat. Meth.*, vol. 5, no. 8, pp. 687–694, 2008.

[173] K. Shafique and M. Shah, "A noniterative greedy algorithm for multiframe point correspondence", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 27, no. 01, pp. 51–65, 2005.

[174] Z. Shaik and V. Asari, "A robust method for multiple face tracking using kalman filter", in *Proc. IEEE Applied Imagery Pattern Recognition Workshop (AIPR'07)*, Washington, DC, USA, Oct. 2007, pp. 125–130.

[175] Y. Sheikh, M. Sheikh, and M. Shah, "Exploring the space of a human action", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'05)*, Beijing, China, Oct. 2005, pp. 144–149.

[176] I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, "Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: application to microtubule growth analysis", *IEEE Trans. Med. Imaging*, vol. 27, no. 6, pp. 789–804, 2008.

[177] I. Smal, M. Loog, W. Niessen, and E. Meijering, "Quantitative comparison of spot detection methods in fluorescence microscopy", *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 282–301, 2010.

[178] I. Smal, E. Meijering, K. Draegestein, N. Galjart, I. Grigoriev, A. Akhmanova, M. E. van Royen, A. B. Houtsmuller, and W. Niessen, "Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering", *Med. Image Anal.*, vol. 12, no. 6, pp. 764–777, 2008.

[179] I. Smal, W. Niessen, and E. Meijering, "A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, May 2008, pp. 264–267.

[180] G. J. D. Smith, D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, J. S. M. Peiris, Y. Guan, and A. Rambaut, "Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic", *Nature*, vol. 459, no. 7250, pp. 1122–1125, 2009.

[181] K. Smith, S. Ba, J.-M. Odobez, and D. Gatica-Perez, "Tracking the visual focus of attention for a varying number of wandering people", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1212–1229, 2008.

[182] L. Song, E. Hennink, I. Young, and H. Tanke, "Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy", *Biophys. J.*, vol. 68, no. 6, pp. 2588–2600, 1995.

[183] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing: Analysis and Machine Vision*. Pacific Grove, CA, USA: Brooks, 1998.

[184] C. Stauffer, "Estimating tracking sources and sinks", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop (CVPRW'03)*, vol. 4, Madison, WI, USA, Jun. 2003, p. 35.

[185] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'99)*, Fort Collins, Colorado, USA, Jun. 1999, pp. 2246–2252.

[186] ——, "Learning patterns of activity using real-time tracking", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.

[187] E. H. K. Stelzer, "Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy", *J. Microsc.*, vol. 189, no. 1, pp. 15–24, 1998.

[188] J. Stover, N. Walker, G. P. Garnett, J. A. Salomon, K. A. Stanecki, P. D. Ghys, N. C. Grassly, R. M. Anderson, and B. Schwartländer, "Can we reverse the HIV/AIDS pandemic with an expanded response?", *Lancet*, vol. 360, no. 9326, pp. 73–77, 2002.

[189] D. Theodorou and U. Suter, "Shape of unperturbed linear polymers: polypropylene", *Macromolecules*, vol. 18, no. 6, pp. 1206–1214, 1985.

[190] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser, "Automatic fluorescent tag detection in 3D with super-resolution:application to the analysis of chromosome movement", *J. Microsc.*, vol. 208, pp. 49–64, Oct. 2002.

[191] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots", *Artificial Intelligence*, vol. 128, no. 1-2, pp. 99–141, 2000.

[192] *Photon counting using photomultiplier tubes*, Hamamatsu Photonics K.K., Iwata City, Shizuoka Pref., Japan, Jul. 2005.

[193] J. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters", *IEEE Trans. Autom. Control*, vol. 27, no. 5, pp. 1054–1065, 1982.

[194] F. Tung, J. S. Zelek, and D. A. Clausi, "Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance", *Image Vision Comput.*, vol. 29, no. 4, pp. 230–240, 2011.

[195] W. Tvarusko, M. Bentele, T. Misteli, R. Rudolf, C. Kaether, D. L. Spector, H. H. Gerdes, and R. Eils, "Time-resolved analysis and visualization of dynamic processes in living cells", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 14, pp. 7950–7955, 1999.

[196] D. Tweed and A. Calway, "Tracking many objects using subordinated condensation", in *Proc. British Machine Vision Conf. (BMVC'02)*, Cardiff, UK, Sep. 2002, pp. 283–292.

[197] P. D. Uchil and W. Mothes, "HIV entry revisited", *Cell*, vol. 137, no. 3, pp. 402–404, 2009.

[198] P. Vallotton, A. Ponti, C. Waterman-Storer, E. Salmon, and G. Danuser, "Recovery, Visualization, and Analysis of Actin and Tubulin Polymer Flow in Live Cells: A Fluorescent Speckle Microscopy Study", *Biophys. J.*, vol. 85, pp. 1289–1306, Aug. 2003.

[199] C. J. Veenman, M. J. T. Reinders, and E. Backer, "Resolving motion correspondence for densely moving points", *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 23, no. 1, pp. 54–72, Jan. 2001.

[200] A. Veeraraghavan, R. Chellappa, and M. Srinivasan, "Shape-and-behavior encoded tracking of bee dances", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 463–476, 2008.

[201] A. Veeraraghavan and A. K. R. Chowdhury, "The function space of an activity", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, Jun. 2006, pp. 959–968.

[202] J. Verestóy and D. Chetverikov, "Experimental comparative evaluation of feature point tracking algorithms", in *Proc. Theoretical Foundations of Computer Vision (TFCV'98)*, Dagstuhl, Germany, Mar. 1998, pp. 167–178.

[203] J. Vermaak, A. Doucet, and P. Pérez, "Maintaining multi-modality through mixture tracking", in *Proc. Int. Conf. on Computer Vision (ICCV'03)*, IEEE, Nice, France, Oct. 2003, pp. 1110–1116.

[204] P. Viola and M. J. Jones, "Robust real-time face detection", *Int. J. Comput. Vision*, vol. 57, pp. 137–154, 2 2004.

[205] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 539–555, 2009.

[206] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition", *Comput. Vision Image Understanding*, vol. 115, no. 2, pp. 224–241, 2011.

[207] G. Welch and G. Bishop, "An introduction to the Kalman filter", University of North Carolina at Chapel Hill, Department of Computer Science, Tech. Rep. TR 95-041, 2007.

[208] S. Wörz, P. Sander, M. Pfannmöller, R. Rieker, S. Joos, G. Mechtersheimer, P. Boukamp, P. Lichter, and K. Rohr, "3D geometry-based quantification of colocalizations in multichannel 3D microscopy images of human soft tissue tumors", *IEEE Trans. Med. Imaging*, vol. 29, no. 8, pp. 1474–1484, 2010.

[209] Z. Wu, T. H. Kunz, and M. Betke, "Efficient track linking methods for track graphs using network-flow and set-cover techniques", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11)*, Colorado Springs, CO, USA, Jun. 2011, pp. 1185–1192.

[210] T. Xiang and S. Gong, "Beyond tracking: modelling activity and understanding behaviour", *Int. J. Comput. Vision*, vol. 67, no. 1, pp. 21–51, 2006.

[211] M. Xu, J. Orwell, and G. Jones, "Tracking football players with multiple cameras", in *Proc. Int. Conf. on Image Processing (ICIP'04)*, Singapore, Oct. 2004, pp. 2909–2912.

[212] X. Xu and B. Li, "Rao-Blackwellised particle filter for tracking with application in visual surveillance", *Proc. IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS'05)*, pp. 17–24, Oct. 2005.

[213] Q. Xue and M. C. Leake, "A novel multiple particle tracking algorithm for noisy in vivo data by minimal path optimization within the spatio-temporal volume", in *Proc. IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI'09)*, Boston, MA, USA, Jun. 2009, pp. 1155–1158.

[214] C. Yang, R. Duraiswami, and L. S. Davis, "Fast multiple object tracking via a hierarchical particle filter", in *Proc. IEEE Int. Conf. on Computer Vision (ICCV'05)*, Beijing, China, Oct. 2005, pp. 212–219.

[215] L. Yang, Z. Qiu, A. Greenaway, and W. Lu, "A new framework for particle detection in low-SNR fluorescence live-cell images and its application for improved particle tracking", *IEEE Trans. Biomed. Eng.*, vol. 59, no. 7, pp. 2040–2050, 2012.

[216] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, 2002.

[217] S. Yang, D. Köhler, K. Teller, T. Cremer, P. L. Baccon, E. Heard, R. Eils, and K. Rohr, "Nonrigid registration of 3-D multichannel microscopy images of cell nuclei", *IEEE Trans. Image. Process.*, vol. 17, no. 4, pp. 493–499, 2008.

[218] J. W. Yoon, A. Bruckbauer, W. J. Fitzgerald, and D. Klenerman, "Bayesian inference for improved single molecule fluorescence tracking", *Biophys. J.*, vol. 94, no. 12, pp. 4932–4947, 2008.

[219] Q. Yu, G. Medioni, and I. Cohen, "Multiple target tracking using spatio-temporal markov chain monte carlo data association", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'07)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[220] T. Yu and Y. Wu, "Collaborative tracking of multiple targets", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2004 (CVPR 2004)*, vol. 1, Washington D. C., USA, Jun. 2004, pp. 834–841.

[221] B. Zhang, J. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for Mixed-Poisson-Gaussian processes and its applications in bioimaging", in *Proc. Int. Conf. on Image Processing (ICIP'07)*, vol. 6, San Antonio, TX, USA, Sep. 2007, pp. 233–236.

[222] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models", *Appl. Opt.*, vol. 46, no. 10, pp. 1819–1829, 2007.

[223] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'08)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[224] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1198–1211, 2008.

[225] W. Zheng and L. Liang, "Fast car detection using image strip features", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'09)*, pp. 2703–2710, Jun. 2009.

[226] S. Zhou, V. Krüger, and R. Chellappa, "Probabilistic recognition of human faces from video", *Comput. Vision Image Understanding*, vol. 91, no. 1-2, pp. 214–245, 2003.

[227] S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters", *IEEE Trans. Image. Process.*, vol. 13, no. 11, pp. 1491–1506, 2004.